

TRƯỜNG ĐẠI HỌC KINH TẾ - ĐẠI HỌC ĐÀ NẴNG

KHOA THƯƠNG MẠI ĐIỆN TỬ



BÀI TẬP NHÓM

HỌC PHẦN: ĐỀ ÁN THỰC HÀNH 1

Đề tài:

Giảng viên	:	TS. Lê Diên Tuấn
Lớp học phần	:	ELC3022_48K29.1
Nhóm	:	Nhóm 2
Sinh viên thực hiện	:	Tiêu Thanh Mai Anh Đình Trần Công Nghĩa Nguyễn Thị Nha

Thành phố Đà Nẵng, năm 2025

MỤC LỤC	
DANH SÁCH THÀNH VIÊN NHÓM.....	4
PHẦN 1: GIỚI THIỆU ĐỀ TÀI	5
1. Lĩnh vực nghiên cứu	5
2. Mục tiêu và phương pháp nghiên cứu.....	5
3. Câu hỏi nghiên cứu	6
4. Các bên liên quan	6
5. Các trường dự kiến cần lấy tại Google Reviews.....	7
5.1. Restaurants.....	7
5.2. Reviews.....	8
PHẦN 2: CƠ SỞ LÝ THUYẾT	10
1. Quy trình nghiên cứu	10
2. Phương pháp cào dữ liệu.....	11
3. Phương pháp gán nhãn	11
3.1. Tổng quan về phương pháp gán nhãn	11
3.2. Thiết lập mô hình phân loại.....	11
3.3. Chuẩn bị dữ liệu huấn luyện.....	12
3.4. Xử lý văn bản - Tokenization.....	12
3.5. Huấn luyện mô hình với PEFT (LoRA).....	13
3.6. Dự đoán và gán nhãn dữ liệu mới	13
3.7. Tổng kết.....	14
PHẦN 3: TRIỂN KHAI DỰ ÁN	14
1. Thu thập dữ liệu:	14
2. Tiền xử lý dữ liệu	16
2.1. Làm sạch dữ liệu.....	16
2.2. Rời rạc hoá dữ liệu.....	18
2.3. Biến đổi dữ liệu	20
PHẦN 4: TRỰC QUAN HOÁ DỮ LIỆU.....	22

1. Sơ đồ quan hệ thực thể.....	22
--------------------------------	----

DANH SÁCH THÀNH VIÊN NHÓM

STT	Họ và tên	Mã số sinh viên
1	Tiêu Thanh Mai Anh	221124029101
2	Đinh Trần Công Nghĩa	221124029127
3	Nguyễn Thị Nha	221124029130

PHẦN 1: GIỚI THIỆU ĐỀ TÀI

1. Lĩnh vực nghiên cứu

Quản trị du lịch nhà hàng và khách sạn.

2. Mục tiêu và phương pháp nghiên cứu

Đà Nẵng là một trong những điểm đến du lịch hấp dẫn nhất Việt Nam, thu hút hàng triệu du khách mỗi năm nhờ bãi biển đẹp, ẩm thực đa dạng và dịch vụ du lịch phát triển. Tuy nhiên, việc hiểu rõ hành vi và sở thích ăn uống của du khách quốc tế tại Đà Nẵng vẫn còn nhiều hạn chế do phương pháp thu thập dữ liệu truyền thống như khảo sát và bảng hỏi có tính tốn kém, mất thời gian và chưa phản ánh đầy đủ trải nghiệm thực tế. Nghiên cứu này nhằm khai thác dữ liệu từ đánh giá trực tuyến trên các nền tảng như **Google Reviews**, sử dụng phương pháp xử lý ngôn ngữ tự nhiên (NLP) và mô hình ngôn ngữ lớn (LLMs - **Large Language Models**) để phân tích sâu hơn về sở thích ẩm thực, mức độ hài lòng, và các yếu tố quan trọng ảnh hưởng đến quyết định ăn uống của du khách.

Ứng dụng của LLMs trong nghiên cứu:

LLMs là một bước tiến vượt bậc trong xử lý ngôn ngữ tự nhiên, có khả năng phân tích dữ liệu văn bản lớn với ngữ cảnh phong phú. Trong nghiên cứu này, LLMs sẽ hỗ trợ:

- Trích xuất thông tin chi tiết từ đánh giá nhà hàng: Nhận diện từ khóa quan trọng, tâm lý du khách, và các yếu tố ảnh hưởng đến trải nghiệm ẩm thực (hương vị, dịch vụ, không gian...).
- Phân loại và nhóm sở thích ăn uống: Xác định các nhóm du khách theo khẩu vị, nhu cầu ăn uống (ví dụ: yêu thích hải sản, ăn chay, fine dining...).
- Cải thiện độ chính xác trong phân tích cảm xúc: So với các phương pháp truyền thống, LLMs có thể hiểu sâu sắc hơn về sắc thái ngôn ngữ, giúp phân biệt những đánh giá tích cực, tiêu cực hoặc trung lập chính xác hơn.
- Tổng hợp dữ liệu quy mô lớn: Với hơn hàng chục nghìn đánh giá, LLMs giúp xử lý dữ liệu khổng lồ một cách nhanh chóng, tiết kiệm chi phí so với phân tích thủ công.

Bằng cách kết hợp dữ liệu lớn, NLP, và LLMs, nghiên cứu này không chỉ giúp nhận diện rõ hơn xu hướng ẩm thực của du khách tại Đà Nẵng, mà còn cung cấp cơ sở khoa học để các nhà quản lý nhà hàng, doanh nghiệp du lịch tối ưu hóa chiến lược kinh doanh và nâng cao trải nghiệm ẩm thực cho khách quốc tế.

→ Phân tích dữ liệu đánh giá nhà hàng trên quy mô lớn, tập trung vào:

- Sở thích ăn uống của du khách.
- Tâm lý & cảm xúc trong đánh giá nhà hàng.

Hệ thống đề xuất nhà hàng cá nhân hóa:

Một hướng tiếp cận quan trọng trong nghiên cứu này là xây dựng **“Hệ thống đề xuất nhà hàng thông minh”**. Mô hình này sử dụng machine learning để đề xuất nhà hàng phù hợp với thông tin cá nhân của khách hàng. Hệ thống không chỉ giúp khách hàng lựa chọn nhà hàng theo sở thích cá nhân mà còn hỗ trợ các công ty du lịch và lữ hành tìm kiếm nhà hàng phù hợp để thiết kế tour du lịch cho khách hàng.

Mục tiêu của hệ thống:

- Tự động đề xuất nhà hàng dựa trên thông tin khách hàng.
- Cung cấp trải nghiệm ăn uống tối ưu bằng cách cá nhân hóa gợi ý.
- Hỗ trợ các doanh nghiệp du lịch, nhà hàng trong việc tiếp cận đúng đối tượng khách hàng.

3. Câu hỏi nghiên cứu

- Những món ăn nào được du khách từ mỗi quốc gia ưa thích?
- Du khách thích ăn gì vào các bữa ăn khác nhau trong ngày?
- Nhận diện điểm mạnh, điểm yếu của mỗi nhà hàng?
 - Xác định những yếu tố khách hàng đánh giá cao (ví dụ: không gian đẹp, giá cả hợp lý).
 - Nhận diện những điểm cần cải thiện (ví dụ: phục vụ chậm, món ăn không đúng khẩu vị).
- Khách hàng thường ưa chuộng những món ăn nào khi đến đây?
- Sở thích về món ăn của du khách có sự thay đổi theo mùa, thời gian hay không?
- Khách hàng thường đến đây một mình, theo cặp, hay gia đình, 1 nhóm lớn?
- Xác định các yếu tố được khách hàng đánh giá cao: món ăn, dịch vụ, giá cả?

4. Các bên liên quan

- **Nhà hàng và chủ sở hữu:**
 - Tối ưu hoá thực đơn và dịch vụ: điều chỉnh menu phù hợp, cải thiện chất lượng dịch vụ giúp tăng trải nghiệm khách hàng.
 - Nắm rõ các xu hướng ẩm thực, sở thích và hành vi ăn uống của khách hàng nhằm xây dựng chiến lược kinh doanh phù hợp, duy trì lợi thế cạnh tranh.

- **Khách hàng:**
 - Giúp họ tìm kiếm được các món ăn, nhà hàng phù hợp với nhu cầu của họ dựa trên những thông tin chính xác và đáng tin cậy từ dữ liệu đánh giá của người dùng thực tế.
 - Cá nhân hoá tìm kiếm và gợi ý dựa trên sở thích, hành vi, xu hướng.
 - Gia tăng trải nghiệm của họ khi sử dụng dịch vụ ẩm thực.
- **Nền tảng đánh giá trực tuyến (Google Reviews):**
 - Nâng cao thuật toán gợi ý nhà hàng dựa trên dữ liệu đánh giá của khách du lịch.
 - Cải thiện hệ thống xếp hạng, giúp người dùng có trải nghiệm tìm kiếm tốt hơn.
- **Công ty du lịch và lữ hành:**
 - Xây dựng các tour du lịch ẩm thực dựa trên sở thích của du khách từ các nền văn hóa khác nhau.

5. Các trường dữ kiện cần lấy tại Google Reviews

5.1. Restaurants

STT	Fields	Data types	Definition	Value
1	Restaurant_id	Int	Mã định danh nhà hàng	
2	Restaurant_Name	Text	Tên nhà hàng	
3	Restaurant_Type	Text	Loại nhà hàng	Vietnamese restaurant, Restaurant, Fine dining restaurant, Steak house, Indian restaurant,...
4	Rating_average	Float	Điểm đánh giá trung bình	
5	Num_of_Reviews	int	Tổng số lượt đánh giá	

6	Price_level	Text	Mức giá	₫100–200K, ₫1,000,000+,....
7	Address	Text	Địa chỉ	
8	Phone	String	Số điện thoại	
9	Open_hour	String	Giờ mở cửa	
10	Accessibility_info	String	Tiện nghi	
11	Created_at	Date	Ngày thu thập dữ liệu	

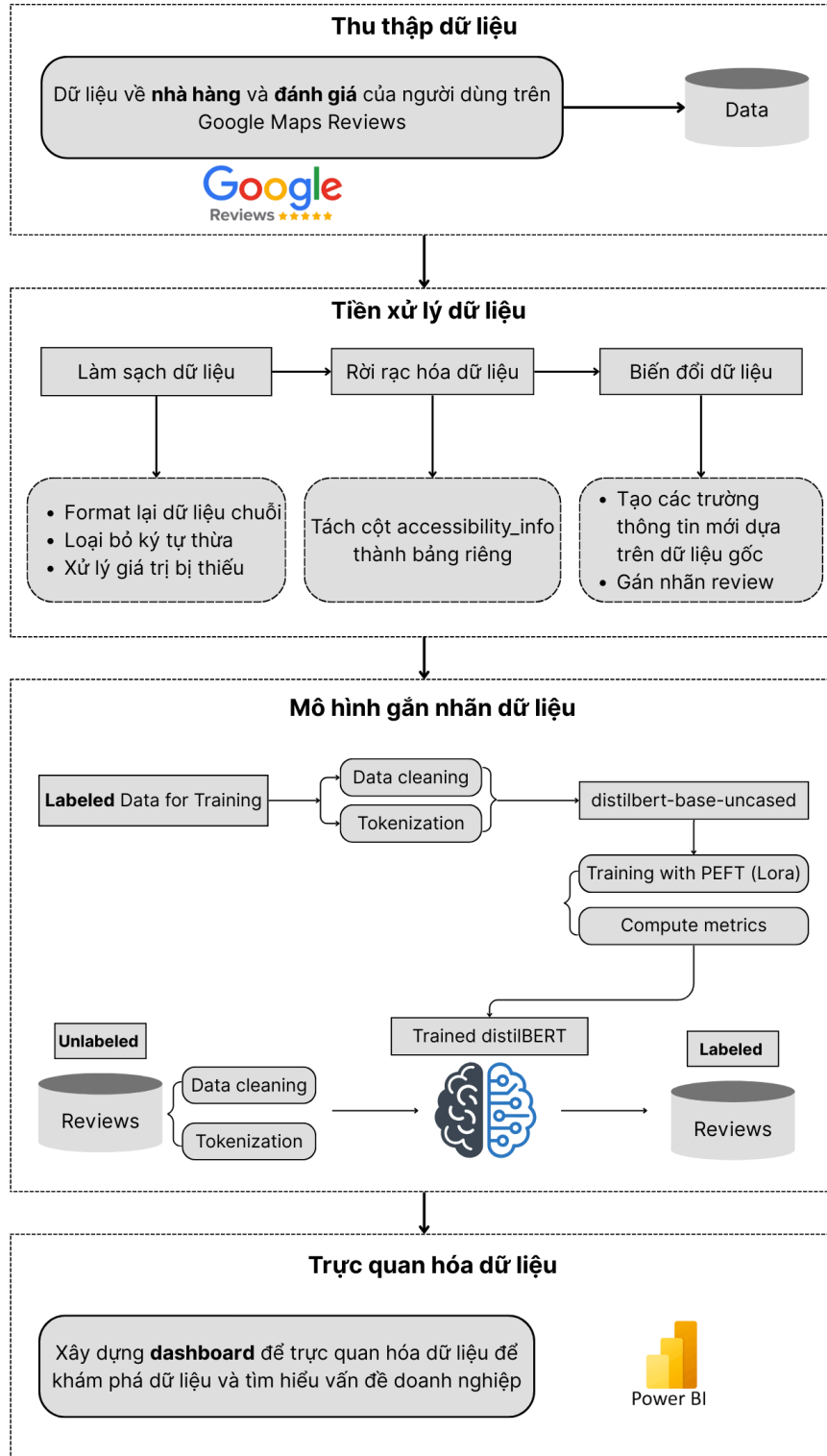
5.2. Reviews

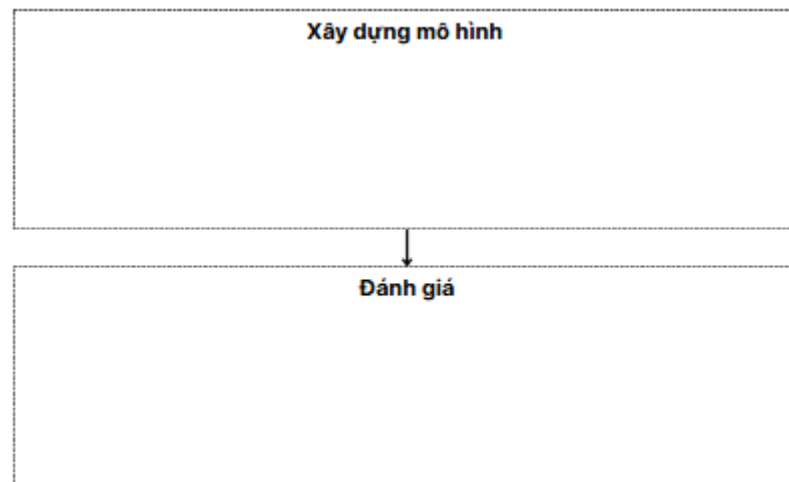
STT	Fields	Data Type	Definition	Value
1	Review_id	Int	Mã định danh cho mỗi review	
2	Restaurant_id	Int	Mã định danh nhà hàng	
3	Reviewer_name	Text	Tên người đánh giá	
4	Reviewer_info	Text	Thông tin về người đánh giá	Gồm số lượng đánh giá đã viết và ảnh đã đăng trên gg review của người đánh giá
5	Rating	Float	Điểm đánh giá	1 → 5
6	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

7	Review_Text	Text	Đánh giá của khách hàng về nhà hàng	
8	Service_rating	Float	Điểm đánh giá dịch vụ	
9	Food_rating	Float	Điểm đánh giá món ăn	
10	Atmosphere_rating	Float	Điểm đánh giá không gian	
11	Service_type	Text	Loại dịch vụ	Dine in
12	Meal_type	Text	Bữa ăn	Breakfast, Lunch, Dinner,...
13	Nation	Text	Quốc tịch người đánh giá	
14	Created_at	Datetime	Ngày thu thập dữ liệu	

PHẦN 2: CƠ SỞ LÝ THUYẾT

1. Quy trình nghiên cứu





2. Phương pháp cào dữ liệu

3. Phương pháp gán nhãn

3.1. Tổng quan về phương pháp gán nhãn

Phương pháp gán nhãn được áp dụng trong nghiên cứu này là phương pháp sử dụng mô hình học sâu tiên tiến nhằm phân loại cảm xúc của các đánh giá văn bản. Mục tiêu của phương pháp là gán nhãn tự động cho các đánh giá với ba nhãn cảm xúc: Negative, Neutral, và Positive. Để đạt được hiệu quả cao trong việc phân loại, chúng tôi áp dụng kỹ thuật tinh chỉnh mô hình ngôn ngữ chuyển đổi (Transformer) thông qua mô hình DistilBERT kết hợp với kỹ thuật Parameter-Efficient Fine-Tuning (PEFT) sử dụng Low-Rank Adaptation (LoRA) nhằm tối ưu hóa khả năng tổng quát hóa và giảm thiểu số lượng tham số cần huấn luyện.

3.2. Thiết lập mô hình phân loại

Khởi tạo và cấu hình mô hình: Mô hình DistilBERT (distilbert-base-uncased) được chọn làm nền tảng cho quá trình gán nhãn do khả năng cân bằng giữa hiệu năng và tốc độ xử lý. Để phục vụ cho bài toán phân loại cảm xúc, mô hình được cấu hình như sau:

- **Ánh xạ nhãn:** Việc ánh xạ nhãn được thực hiện thông qua hai từ điển id2label và label2id, đảm bảo tính nhất quán giữa các nhãn số (0, 1, 2) và nhãn văn bản (Negative, Neutral, Positive). Điều này hỗ trợ trực tiếp cho quá trình chuyển đổi giữa kết quả mô hình và nhãn thực tế.
- **Cấu hình mô hình:**
 - Sử dụng AutoConfig để tạo cấu hình tùy chỉnh với các tham số như:
 - Số nhãn (num_labels): 3 tương ứng với ba lớp cảm xúc.
 - Tăng tỷ lệ dropout:

- `hidden_dropout_prob: 0.3`
- `attention_probs_dropout_prob: 0.3`
 - Việc tăng tỷ lệ dropout giúp giảm thiểu hiện tượng overfitting, đảm bảo khả năng tổng quát hóa của mô hình.
- Tải mô hình: Khởi tạo mô hình phân loại chuỗi sử dụng “AutoModelForSequenceClassification”, kết hợp với cấu hình đã thiết lập để đảm bảo tính tương thích.

3.3. Chuẩn bị dữ liệu huấn luyện

Tải và làm sạch dữ liệu:

Dữ liệu huấn luyện được lấy từ tập `yqzheng/semeval2014_restaurants`, chứa các đánh giá về nhà hàng với nhãn cảm xúc gốc thuộc tập giá trị `{-1, 0, 1}`. Để phù hợp với yêu cầu của mô hình, các bước tiền xử lý được thực hiện như sau:

- Tải dữ liệu: Sử dụng hàm `load_dataset` từ thư viện `datasets` để tải dữ liệu từ nguồn công khai.
- Chuyển đổi nhãn: Các nhãn gốc được chuyển đổi như sau:
 - `-1` → 0 (Negative)
 - `0` → 1 (Neutral)
 - `1` → 2 (Positive)

3.4. Xử lý văn bản - Tokenization

3.4.1. Khởi tạo và cấu hình tokenizer:

Quá trình xử lý văn bản được thực hiện với `AutoTokenizer` từ checkpoint `distilbert-base-uncased`, đảm bảo sự tương thích với mô hình huấn luyện. Các đặc điểm chính bao gồm:

- Cắt bớt từ trái: Đảm bảo mô hình không bị quá tải bởi các câu quá dài.
- Độ dài tối đa: 512 token, đảm bảo tính toàn vẹn ngữ nghĩa.
- Token đặc biệt: Nếu thiếu `pad_token`, hệ thống sẽ tự động thêm vào và điều chỉnh kích thước embedding tương ứng.

3.4.2. Áp dụng tokenization:

Quá trình token hóa được áp dụng cho cả tập huấn luyện và tập kiểm tra bằng cách sử dụng phương pháp `map` để đảm bảo tính nhất quán.

3.5. Huấn luyện mô hình với PEFT (LoRA)

3.5.1. Kỹ thuật LoRA

LoRA (Low-Rank Adaptation) là một kỹ thuật Parameter-Efficient Fine-Tuning, giúp giảm thiểu số lượng tham số cần tinh chỉnh mà không làm suy giảm hiệu năng của mô hình. Bằng cách thêm các ma trận xếp hạng thấp vào các tham số gốc, LoRA cho phép mô hình học tập trung vào các đặc trưng quan trọng mà không cần cập nhật toàn bộ trọng số.

- Cấu hình LoRA:
 - Rank (r): 4
 - Alpha (lora_alpha): 32
 - Dropout (lora_dropout): 0.01
 - Tầng áp dụng: q_lin (query layer)
- Áp dụng LoRA: Sử dụng hàm `get_peft_model` để kết hợp LoRA với mô hình gốc. Số lượng tham số cần tinh chỉnh giảm xuống còn khoảng 0.9% tổng số tham số ban đầu, giảm đáng kể độ phức tạp tính toán.

3.5.2. Thiết lập huấn luyện

- Tốc độ học: $2e-5$, tối ưu hóa bằng AdamW.
- Kích thước batch: 16, nhằm tận dụng tối đa bộ nhớ GPU.
- Số epoch: 10, đảm bảo hội tụ tốt mà không gây overfitting.
- Chiến lược đánh giá và lưu mô hình:
 - `evaluation_strategy`: epoch
 - `save_strategy`: epoch

Huấn luyện được thực hiện thông qua đối tượng Trainer, kết hợp với hàm tính toán độ chính xác (`compute_metrics`) nhằm đánh giá hiệu năng sau mỗi epoch.

3.6. Dự đoán và gán nhãn dữ liệu mới

Sau khi huấn luyện, mô hình được sử dụng để dự đoán cảm xúc trên tập dữ liệu mới chưa được gán nhãn. Các bước được thực hiện như sau:

- Dự đoán nhãn cảm xúc:
 - Văn bản đầu vào được token hóa với độ dài tối đa 512 token.
 - Mô hình dự đoán logits và nhãn được xác định bằng cách lấy xác suất cao nhất từ các lớp.
- Xuất kết quả:

- Các dự đoán được ánh xạ lại thành nhãn cảm xúc (Negative, Neutral, Positive).
- Dữ liệu đã gán nhãn được lưu trữ dưới dạng file CSV.

3.7. Tổng kết

Phương pháp gán nhãn tự động được triển khai thông qua mô hình DistilBERT kết hợp kỹ thuật LoRA nhằm tối ưu hóa hiệu năng và tốc độ huấn luyện. Quá trình huấn luyện và dự đoán được thực hiện một cách nhất quán và có hệ thống, đảm bảo khả năng gán nhãn chính xác và nhất quán cho các tập dữ liệu mới.

PHẦN 3: TRIỂN KHAI DỰ ÁN

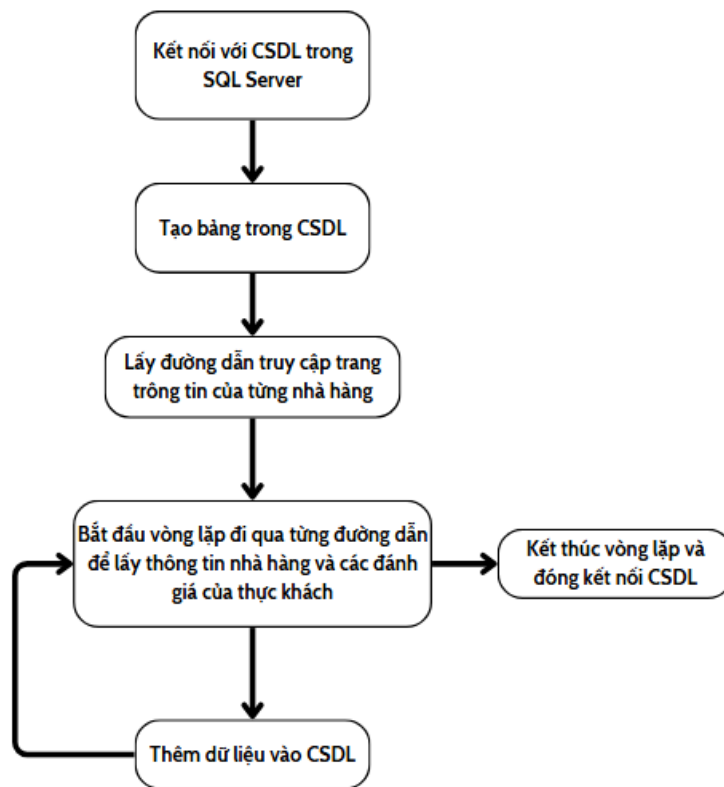
1. Thu thập dữ liệu:

Thu thập dữ liệu là bước đầu tiên khi ta thực hiện bất kỳ 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:



Kết quả:

Bảng Restaurants: Gồm 51 dòng và 11 cột

restaurant_id	restaurant_name	restaurant_type	rating	num_of_reviews	price_level	address	phone	opening_hours	accessibility_info	created_at
1	Nhà hàng NHÀ BẾP CHỢ HÂN	Vietnamese restaurant	4.7	4406	\$100-200K	22 Hùng Vương, Hải Châu 1, Hải Châu, Đà Nẵng 550000	0236 3966 268	Thursday 9 AM-10 PM Friday 9 AM-10 PM Saturday 9 AM-10 PM Sunday 9 AM-10 PM	[No wheelchair-accessible car park], "No wheelchair-accessible seating", "Does not have assistive hearing loop", "No wheelchair-accessible toilet", "Has outdoor seating"	2025-03-20 23:52:16.330
2	Thị Gõ Restaurant Da Nang	Vietnamese restaurant	4.8	6339	\$100-200K	53 Phan Thúc Duyệt, Bắc Mỹ Phú, Ngũ Hành Sơn, Đà Nẵng 550000	Unknown	Friday 10 AM-10 PM Saturday 10 AM-10 PM Sunday 10 AM-10 PM	[Has wheelchair-accessible seating], "Does not have assistive hearing loop", "No wheelchair-accessible toilet", "Has outdoor seating"	2025-03-21 00:31:56.500
3	Bếp Cuốn Đà Nẵng	Vietnamese restaurant	4.8	7430	\$100-300K	31-33 Trần Bạch Đằng, Phước Mỹ, Sơn Trà, Đà Nẵng 550000	Unknown	Friday 10:30 AM-9 PM Saturday 10:30 AM-9 PM Sunday 10:30 AM-9 PM	[Has wheelchair-accessible seating], "Does not have assistive hearing loop", "No wheelchair-accessible toilet", "Has outdoor seating"	2025-03-21 01:14:41.647
4	Cà Ba Phở bò	Restaurant	4.8	5432	\$100-200K	154 Bạch Đằng, Hải Châu 1, Hải Châu, Đà Nẵng 550000	0236 3996 268	Friday 10 AM-10 PM Saturday 10 AM-10 PM Sunday 10 AM-10 PM	[No wheelchair-accessible car park], "No wheelchair-accessible seating", "Does not have assistive hearing loop", "No wheelchair-accessible toilet", "Has outdoor seating"	2025-03-21 02:05:42.390
5	Nhà Hàng Nhà Bếp Khuê Mỹ	Vietnamese restaurant	4.8	4290	\$100-200K	416 Võ Nguyên Giáp, Bắc Mỹ An, Ngũ Hành Sơn, Đà Nẵng 550000	0236 3956 080	Friday 10 AM-10 PM Saturday 10 AM-10 PM Sunday 10 AM-10 PM	[No wheelchair-accessible car park], "No wheelchair-accessible seating", "Does not have assistive hearing loop", "No wheelchair-accessible toilet", "Has outdoor seating"	2025-03-21 02:44:52.247
6	Sky View Restaurant	Fine dining restaurant	4.8	650	Unknown	216 Võ Nguyên Giáp, Phước Mỹ, Sơn Trà, Đà Nẵng 550000	Unknown	Unknown	[Has wheelchair-accessible entrance], "Has wheelchair-accessible toilet", "Has outdoor seating"	2025-03-21 03:26:35.977
7	Ăn Thôi Restaurant	Vietnamese restaurant	4.6	2636	\$100-200K	114 Bạch Đằng, Hải Châu 1, Hải Châu, Đà Nẵng 550000	Unknown	Friday 10:30 AM-9:30 PM Saturday 10:30 AM-9:30 PM Sunday 10:30 AM-9:30 PM	[Has wheelchair-accessible toilet], "No wheelchair-accessible car park", "Has outdoor seating"	2025-03-21 04:17:22.507
8	Red Sky Steakhouse	Steak house	4.6	925	\$\$\$	248 Trần Phú, Phước Ninh, Hải Châu, Đà Nẵng 550000	Unknown	Friday 11 AM-2 PM 5-11 PM Saturday 11 AM-2 PM Sunday 11 AM-2 PM	[No wheelchair-accessible car park], "Has no outdoor seating"	2025-03-21 04:56:36.617
9	MAAZI Da Nang	Indian restaurant	4.8	2683	\$100-200K	264 Trần Phú, Phước Ninh, Hải Châu, Đà Nẵng 700000	Unknown	Friday 11 AM-12 AM Saturday 11 AM-12 AM Sunday 11 AM-12 AM	[Has wheelchair-accessible seating], "Has wheelchair-accessible car park", "Has outdoor seating"	2025-03-21 06:01:59.347
10	Redcrab Danang	Traditional restaurant	4.8	1319	Unknown	138 Phạm Văn Đồng, street, Sơn Trà, Đà Nẵng 550000	0934 724 919	Friday 10 AM-10 PM Saturday 10 AM-10 PM Sunday 10 AM-10 PM	[Has wheelchair-accessible car park], "Has wheelchair-accessible seating", "No wheelchair-accessible toilet", "Has outdoor seating"	2025-03-21 15:09:43.847
11	Nhà Bếp Kìa Restaurant	Vietnamese restaurant	4.8	5819	\$100-200K	64B Hà Bổng, Phước Mỹ, Sơn Trà, Đà Nẵng 550000	0905 123 858	Friday 10 AM-9:30 PM Saturday 10 AM-9:30 PM Sunday 10 AM-9:30 PM	[Has wheelchair-accessible seating], "No wheelchair-accessible car park", "Has outdoor seating"	2025-03-21 15:30:05.413
12	LoCo Restaurant - Danang seafood restaurant - CH	Restaurant	4.8	2288	\$\$\$	1C D. Lê Duẩn, P. Hải Châu, Đà Nẵng 550000	Unknown	Friday 10:30 AM-12 AM Saturday 10:30 AM-12 AM Sunday 10:30 AM-12 AM	[Has wheelchair-accessible toilet], "Has outdoor seating"	2025-03-21 16:09:03.457
13	Nhà Hàng Sĩ Dining	Italian restaurant	4.9	461	\$1,000,000+	1 Giang Châu 2 Phường Khuê Mỹ, Quận Ngũ Hành Sơn, Đà Nẵng 550000	0774 464 441	"Friday": "5:30-9:30 PM", "Saturday": "5:30-9:30 PM", "Sunday": "5:30-9:30 PM"	[Has wheelchair-accessible entrance], "Has wheelchair-accessible car park", "Has outdoor seating"	2025-03-21 20:04:39.537
14	Nhà hàng Lân Gió	Vietnamese restaurant	4.9	1066	Unknown	169 Trần Phú, Hải Châu 1, Hải Châu, Đà Nẵng 550000	Unknown	"Friday": "10:30 AM-10:30 PM", "Saturday": "10:30 AM-10:30 PM", "Sunday": "10:30 AM-10:30 PM"	[Offers takeaway], "Serves dine-in", "Popular for lunch and dinner", "Has outdoor seating"	2025-03-21 22:30:40.427
15	Nhà hàng Thiên Kim	Restaurant	4.8	6136	\$100-300K	166 Bạch Đằng, Hải Châu 1, Hải Châu, Đà Nẵng 550000	Unknown	"Friday": "10:30 AM-10:30 PM", "Saturday": "10:30 AM-10:30 PM", "Sunday": "10:30 AM-10:30 PM"	[Has wheelchair-accessible entrance], "Has wheelchair-accessible car park", "Has outdoor seating"	2025-03-21 22:59:30.207
17	MaOON Siro Danang	Restaurant	4.6	308	Unknown	4A Đào Xanh 4, Khu Biệt Thự Đào Xanh, Hải Châu, Đà Nẵng 550000	0905 263 891	"Saturday": "7 AM-10 PM", "Sunday": "9-10 PM", "Monday": "9-10 PM"	[Does not have assistive hearing loop], "No wheelchair-accessible toilet", "Has outdoor seating"	2025-03-22 12:26:23.967
18	SIX ON SIX CAFE - BRUNCH & RESTAURANT	Restaurant	4.8	1664	\$100-200K	64 Bà Huyện Thanh Quan, Bắc Mỹ An, Ngũ Hành Sơn, Đà Nẵng 550000	Unknown	"Saturday": "7 AM-10 PM", "Sunday": "9-10 PM", "Monday": "9-10 PM"	[Has wheelchair-accessible entrance], "Has wheelchair-accessible car park", "Has outdoor seating"	2025-03-22 12:59:49.923
19	Mỹ Quăng Cà Sầu	Restaurant	4.7	880	\$1-100,000	397 B. Trần Hưng Đạo, An Hải Trung, Sơn Trà, Đà Nẵng 550000	Unknown	"Saturday": "6 AM-11 PM", "Sunday": "6 AM-11 PM", "Monday": "6 AM-11 PM"	[Has wheelchair-accessible car park], "Has wheelchair-accessible seating", "Has outdoor seating"	2025-03-22 13:44:00.450
20	EMO'S HOME COOKED VIETNAMESE CUISINE	Vietnamese restaurant	4.8	2232	\$100-200K	Đường Mỹ Khê 4, Phước Mỹ, Sơn Trà, Đà Nẵng 550000	0934 844 988	"Saturday": "6 AM-11 PM", "Sunday": "6 AM-11 PM", "Monday": "6 AM-11 PM"	[Has wheelchair-accessible seating], "Has wheelchair-accessible car park", "Has outdoor seating"	2025-03-22 14:30:42.140
21	Restaurant PHO HOUSE	Restaurant	4.7	413	\$100-200K	111 Nại Nam 5, Hoà Cường Bắc, Hải Châu, Đà Nẵng 550000	0708 352 141	"Saturday": "10 AM-10 PM", "Sunday": "10 AM-10 PM", "Monday": "10 AM-10 PM"	[No wheelchair-accessible car park], "Offers takeaway", "Has outdoor seating"	2025-03-22 15:12:30.183
22	Nhà Hàng Chay Ngọc Chi - Vegetarian Restaurant M	Vegetarian restaurant	4.6	971	\$1-100,000	202 Hoàng Diệu, Nam Dương, Hải Châu, Đà Nẵng 550000	0935 085 088	"Saturday": "7 AM-9 PM", "Sunday": "7 AM-9 PM", "Monday": "7 AM-9 PM"	[Has wheelchair-accessible entrance], "Has wheelchair-accessible car park", "Has outdoor seating"	2025-03-22 15:31:34.860
23	Anchor Beer Club	Restaurant	4.4	342	\$\$\$	29 Quang Đông, Vĩnh Trung, Thanh Khê, Đà Nẵng 550000	Unknown	"Saturday": "10 AM-12 AM", "Sunday": "10 AM-12 AM", "Monday": "10 AM-12 AM"	[No wheelchair-accessible car park], "No wheelchair-accessible seating", "Has outdoor seating"	2025-03-22 16:26:13.847
24	Nhà Hàng Đà Nẵng - Ho Xanh Restaurant	Restaurant	4	177	\$\$\$	K 1 Đỗ Quang, Vĩnh Trung, Thanh Khê, Đà Nẵng 550000	0905 718 477	"Saturday": "9 AM-10:30 PM", "Sunday": "9 AM-10:30 PM", "Monday": "9 AM-10:30 PM"	[Has wheelchair-accessible seating], "Has wheelchair-accessible car park", "Has outdoor seating"	2025-03-22 21:11:12.577
25	Benaras Central Da Nang	Indian restaurant	4.1	202	Unknown	207 Nguyễn Văn Thoại, Anh Hải Đông, Sơn Trà, Đà Nẵng 550000	0961 888 597	"Saturday": "11 AM-11 PM", "Sunday": "11 AM-11 PM", "Monday": "11 AM-11 PM"	[Has wheelchair-accessible seating], "Offers takeaway", "Has outdoor seating"	2025-03-22 22:19:32.750
26	Takoyaki Hatachi Da Nang	Takoyaki restaurant	4.9	64	\$1-100,000	118 Phan Thanh, Thạc Gián, Thanh Khê, Đà Nẵng 550000	0917 012 369	"Saturday": "4-10 PM", "Sunday": "4-10 PM", "Monday": "4-10 PM"	[Offers delivery], "Offers takeaway", "Serves dine-in", "Has outdoor seating"	2025-03-22 23:16:43.770
27	La Rendee Vous	French restaurant	4.3	559	\$1,000,000+	28 Nguyễn Cao Lược, An Hải Bắc, Sơn Trà, Đà Nẵng 550000	Unknown	"Sunday": "11 AM-10 PM", "Monday": "11 AM-10 PM", "Tuesday": "11 AM-10 PM"	[No wheelchair-accessible car park], "No wheelchair-accessible seating", "Has outdoor seating"	2025-03-23 09:25:25.310
28	Bún Bò Bà Risi	Noodle shop	4.5	199	\$1-100,000	5 Phan Thanh Tài, Hòa Thuận Đông, Hải Châu, Đà Nẵng 550000	0704 196 497	"Sunday": "6 AM-10 PM", "Monday": "6 AM-10 PM", "Tuesday": "6 AM-10 PM"	[Offers delivery], "Offers takeaway", "Serves dine-in", "Has outdoor seating"	2025-03-23 09:27:47.143
29	Quán Tỳ	Vietnamese restaurant	4.3	87	\$1-100,000	5 An Thượng 6, Bắc Mỹ Phú, Ngũ Hành Sơn, Đà Nẵng 550000	0979 298 398	"Sunday": "6 AM-9 PM", "Monday": "6 AM-9 PM", "Tuesday": "6 AM-9 PM"	[Has wheelchair-accessible seating], "Has wheelchair-accessible car park", "Has outdoor seating"	2025-03-23 10:53:25.670
30	Ăn Thục Nấm Hân - Bánh Xèo Tôm Nhảy	Vietnamese restaurant	4	852	\$1-100,000	48 Phan Thanh, Thạc Gián, Thanh Khê, Đà Nẵng 550000	Unknown	"Sunday": "6 AM-9 PM", "Monday": "6 AM-9 PM", "Tuesday": "6 AM-9 PM"	[Has wheelchair-accessible seating], "Has wheelchair-accessible car park", "Has outdoor seating"	2025-03-23 10:10:29.710

Query executed successfully.

LAPTOP-IP1RMDOK\DATA SERVER... sa (99) - restaurants 00:00:00 51 row

Bảng review: Gồm 14715 dòng và 14 cột

	review_id	restaurant_id	reviewer_name	reviewer_info	rating	review_time	review_text	service_rating	food_rating	atmosphere_rating	service_type	meal_type	nation	created_at
1	Rv00001	1	Joe Greco	Local Guide · 46 reviews · 28 photos	5	3 weeks ago	As soon as the food arrived I knew I'd made the right ...	5	5	5	Unknown	Unknown	Unknown	2025-03-20 23:57:54.067
2	Rv00002	1	Annie Song	Local Guide · 172 reviews · 684 photos	5	3 weeks ago	As Vietnamese living overseas, we loved the flavours ...	5	5	5	Dine in	Lunch	Unknown	2025-03-20 23:57:43.930
3	Rv00003	1	Jiwon Kim	4 reviews · 11 photos	5	2 months ago	It was a huge and cozy place to try Vietnam style food!...	5	5	5	Dine in	Lunch	Unknown	2025-03-20 23:58:04.227
4	Rv00004	1	katie tang	Local Guide · 11 reviews · 14 photos	3	5 months ago	Had the beef pho and fried rice. Flavors not too authe...	3	3	4	Dine in	Lunch	Unknown	2025-03-21 00:00:24.627
5	Rv00005	1	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 this pla...	5	5	5	Unknown	Unknown	Unknown	2025-03-21 00:00:16.657
6	Rv00006	1	Richard Hinchman	Local Guide · 83 reviews · 313 photos	3	2 months ago	Very Koreanized version of Vietnamese food. We cam...	3	1	5	Unknown	Unknown	Unknown	2025-03-20 23:58:34.717
7	Rv00007	1	Eli T	Local Guide · 19 reviews · 11 photos	4	3 months ago	Tourist-friendly place to eat. Clean and accommodati...	5	4	5	Dine in	Lunch	Unknown	2025-03-20 23:58:44.907
8	Rv00008	1	Emily Lee	Local Guide · 47 reviews · 204 photos	4	3 months ago	Service is good. Staff are friendly and attentive. The fo...	4	4	4	Dine in	Lunch	Unknown	2025-03-20 23:59:15.510
9	Rv00009	1	Erng Mitphumwiboon	Local Guide · 11 reviews · 25 photos	5	2 years ago	Foods are nice and the portion is generous. We order...	N/A	N/A	N/A	Dine in	Dinner	Unknown	2025-03-21 00:02:49.523
10	Rv00010	1	ThanhHa H.	Local Guide · 123 reviews · 1,222 photos	3	a month ago	Located across from Cho Hàn... very busy at lunch tim...	4	3	4	Dine in	Lunch	Unknown	2025-03-20 23:59:05.350
11	Rv00011	1	Boon Keng	Local Guide · 130 reviews · 229 photos	5	3 months ago	Restaurant was huge and it was still full during week...	5	5	5	Dine in	Dinner	Unknown	2025-03-20 23:59:25.777
12	Rv00012	1	Elconore A.	Local Guide · 49 reviews · 301 photos	4	4 months ago	We took the 9 dishes menu set for 2 at 210,000 dong!...	4	4	4	Dine in	Dinner	Unknown	2025-03-21 00:00:08.557
13	Rv00013	1	Bernard Lee	1 review	5	2 months ago	Large variety of choices, from set menus to ala carte...	5	5	4	Dine in	Lunch	Unknown	2025-03-21 00:08:03.887
14	Rv00014	1	Leslie Predy	2 reviews · 1 photo	5	2 months ago	Very busy, but so efficiently run, we didn't have to wait...	5	5	5	Unknown	Unknown	Unknown	2025-03-21 00:00:57.317
15	Rv00015	1	Gloria LEE	2 reviews · 8 photos	4	2 weeks ago	Nice environment, good air-conditioning and good se...	4	4	4	Dine in	Lunch	Unknown	2025-03-21 00:01:38.030
16	Rv00016	1	christina lea ekawaty	Local Guide · 146 reviews · 105 photos	5	a year ago	Nice place to try authentic vietnamese food. Good am...	5	5	5	Dine in	Lunch	Unknown	2025-03-21 00:01:27.840
17	Rv00017	1	ZB Liang	7 reviews · 3 photos	5	3 months ago	We bumped into this place after a visit to Han Market...	5	5	5	Dine in	Lunch	Unknown	2025-03-21 00:09:25.010
18	Rv00018	1	Justin Phua	Local Guide · 55 reviews · 6 photos	5	2 months ago	The pho featured a clear, flavorful broth and tender b...	5	5	5	Dine in	Dinner	Unknown	2025-03-21 00:01:07.460
19	Rv00019	1	Jack Lim	Local Guide · 219 reviews · 274 photos	5	5 months ago	Great place, good food, and fair price. We enjoy many...	5	4	5	Dine in	Lunch	Unknown	2025-03-21 00:01:58.483
20	Rv00020	1	Kelly	Local Guide · 141 reviews · 685 photos	5	11 months ago	One of my favorite restaurants in DaNang. The Hoi A...	5	5	5	Dine in	Lunch	Unknown	2025-03-21 00:02:18.963
21	Rv00021	1	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	2025-03-21 00:09:55.333
22	Rv00022	1	Hup Lee Tan	Local Guide · 19 reviews · 13 photos	5	3 months ago	Good food and great service. One of the staff even sh...	5	5	5	Dine in	Lunch	Unknown	2025-03-21 00:01:48.283
23	Rv00023	1	Boy Truong	14 reviews · 9 photos	1	5 months ago	Due to the high reviews, our expectations were set ve...	3	3	4	Unknown	Unknown	Unknown	2025-03-21 00:03:09.807
24	Rv00024	1	Lydia See	2 reviews · 6 photos	5	3 weeks ago	Great food and affordable price! One of the best tasti...	5	4	5	Unknown	Dinner	Unknown	2025-03-21 00:02:59.710
25	Rv00025	1	Michael Huano	1 review	5	2 weeks ago	good food and free GREAT mango smoothie ask for ...	N/A	N/A	N/A	Unknown	Unknown	Unknown	2025-03-21 00:11:26.580

Query executed successfully.

LAPTOP-IP1RMDOK\DATASERVERN... sa (144) restaurants 000002 14,715 rows

2. Tiền xử lý dữ liệu

2.1. Làm sạch dữ liệu

a. Loại bỏ ký tự thừa, không hợp lệ trong chuỗi:

Dữ liệu cột “Opening_hours” trong quá trình chuyển đổi nội dung từ web vào CSDL trong SQL Server bị lỗi encoding, khiến cho nội dung cào về chứa cả những ký tự đặc biệt dưới dạng mã Unicode.

	id	url	title	opening_hours
1	1	https://w...	Nhà hàng NHÀ BẾP CHỢ HÀN	Thursday 9 AM-10 PM Friday 9 AM-10 PM Saturday 9 AM-10 PM Sunday 9 AM-10 PM Monday 9 AM-10 PM Tuesday 9 AM-10 PM Wednesday 9 AM-10 PM Su...
2	2	https://w...	Thia Gổ Restaurant Da Nang	Friday 10 AM-10 PM Saturday 10 AM-10 PM Sunday 10 AM-10 PM Monday 10 AM-10 PM Tuesday 10 AM-10 PM Wednesday 10 AM-10 PM Thursday 10 AM-1...
3	3	https://w...	Bếp Cườn Đà Nẵng	Friday 10:30 AM-9 PM Saturday 10:30 AM-9 PM Sunday 10:30 AM-9 PM Monday 10:30 AM-9 PM Tuesday 10:30 AM-9 PM Wednesday 10:30 AM-9 PM Thursd...
4	4	https://w...	Cô Ba Phở bò	Friday 10 AM-10 PM Saturday 10 AM-10 PM Sunday 10 AM-10 PM Monday 10 AM-10 PM Tuesday 10 AM-10 PM Wednesday 10 AM-10 PM Thursday 10 AM-1...
5	5	https://w...	Nhà Hàng Nhà Bếp Khuê Mỹ	Friday 10 AM-10 PM Saturday 10 AM-10 PM Sunday 10 AM-10 PM Monday 10 AM-10 PM Tuesday 10 AM-10 PM Wednesday 10 AM-10 PM Thursday 10 AM-1...
6	6	https://w...	Sky View Restaurant	
7	7	https://w...	Ăn Thôi Restaurant	Friday 10:30 AM-9:30 PM Saturday 10:30 AM-9:30 PM Sunday 10:30 AM-9:30 PM Monday 10:30 AM-9:30 PM Tuesday 10:30 AM-9:30 PM Wednesday 10:30 A...
8	8	https://w...	Red Sky Steakhouse	Friday 11 AM-2 PM 5-11 PM Saturday 11 AM-2 PM 5-11 PM Sunday 11 AM-2 PM 5-11 PM Monday 11 AM-2 PM 5-11 PM Tuesday 11 AM-2 PM 5-11 PM Wedn...
9	10	https://w...	MAAZI Da Nang	Friday 11 AM-12 AM Saturday 11 AM-12 AM Sunday 11 AM-12 AM Monday 11 AM-12 AM Tuesday 11 AM-12 AM Wednesday 11 AM-12 AM Thursday 11 AM-12...
21	21	https://w...	Redcrab Danang	Friday 10 AM-10 PM Saturday 10 AM-10 PM Sunday 10 AM-10 PM Monday 10 AM-10 PM Tuesday 10 AM-10 PM Wednesday 10 AM-10 PM Thursday 10 AM-1...
11	23	https://w...	Nhà Bếp Xua Restaurant	Friday 10 AM-9:30 PM Saturday 10 AM-9:30 PM Sunday 10 AM-8 PM Monday 10 AM-9:30 PM Tuesday 10 AM-9:30 PM Wednesday 10 AM-9:30 PM Thursday ...
12	24	https://w...	LoCo Restaurant - Danang seafoo...	Friday 10:30 AM-12 AM Saturday 10:30 AM-12 AM Sunday 10:30 AM-12 AM Monday 10:30 AM-12 AM Tuesday 10:30 AM-12 AM Wednesday 10:30 AM-12 AM ...
13	27	https://w...	Nhà Hàng Sĩ Dining	("Friday": "5:30u20139:30u202fPM", "Saturday": "5:30u20139:30u202fPM", "Sunday": "5:30u20139:30u202fPM", "Monday": "5:30u20139:30u202fPM", "T...
14	28	https://w...	Gypsy Rooftop Restaurant & Bar	("Friday": "5:30u20139:30u202fPM", "Saturday": "5:30u20139:30u202fPM", "Sunday": "5:30u20139:30u202fPM", "Monday": "5:30u20139:30u202fPM", "T...
15	29	https://w...	Nhà hàng Lân Gió	("Friday": "10:30u202fAMu201310:30u202fPM", "Saturday": "10:30u202fAMu201310:30u202fPM", "Sunday": "10:30u202fAMu201310:30u202fPM", "Mo...
16	30	https://w...	Nhà hàng Thiên Kim	("Friday": "10:30u202fAMu201310:30u202fPM", "Saturday": "10:30u202fAMu201310:30u202fPM", "Sunday": "10:30u202fAMu201310:30u202fPM", "Mo...
17	32	https://w...	MeiOh Bistro Danang	("Saturday": "7u202fAMu201310u202fPM", "Sunday": "3u201310u202fPM", "Monday": "3u201311u202fPM", "Tuesday": "3u201310u202fPM", "Wednes...

Trong quá trình xử lý, nhóm nhận thấy lỗi ký tự mã unicode được hiểu như sau:

- \u2013 tương ứng với dấu “-”
- \u202f tương ứng với dấu “ ”

Do đó nhóm thực hiện thay thế các mã ký tự unicode thành giá trị đúng của nó bằng SQL Server. Kết quả:

	uri	title	opening_hours
1	https://www...	Nhà hàng NHÀ BẾP CHỢ HÀN	Thursday 9AM-10PM Friday 9AM-10PM Saturday 9AM-10PM Sunday 9AM-10PM Monday 9AM-...
2	https://www...	Redcrab Danang	Friday 10AM-10PM Saturday 10AM-10PM Sunday 10AM-10PM Monday 10AM-10PM Tuesday 1...
3	https://www...	Nhà Bếp Xua Restaurant	Friday 10AM-9:30PM Saturday 10AM-9:30PM Sunday 10AM-8PM Monday 10AM-9:30PM Tuesda...
4	https://www...	LoCo Restaurant - Danang seafood restaurant - 다낭 ...	Friday 10:30AM-12AM Saturday 10:30AM-12AM Sunday 10:30AM-12AM Monday 10:30AM-12AM ...
5	https://www...	Nhà Hàng Sĩ Dining	("Friday": "5:30-9:30 PM", "Saturday": "5:30-9:30 PM", "Sunday": "5:30-9:30 PM", "Monday": "5:30-9:30 PM...
6	https://www...	Gypsy Rooftop Restaurant & Bar	("Friday": "5:30-9:30 PM", "Saturday": "5:30-9:30 PM", "Sunday": "5:30-9:30 PM", "Monday": "5:30-9:30 PM...
7	https://www...	Nhà hàng Làn Gió	("Friday": "10:30 AM-10:30 PM", "Saturday": "10:30 AM-10:30 PM", "Sunday": "10:30 AM-10:30 PM", "Mon...
8	https://www...	Nhà hàng Thiên Kim	("Friday": "10:30 AM-10:30 PM", "Saturday": "10:30 AM-10:30 PM", "Sunday": "10:30 AM-10:30 PM", "Mon...
9	https://www...	MeliOh Bistro Danang	("Saturday": "7 AM-10 PM", "Sunday": "3-10 PM", "Monday": "3-11 PM", "Tuesday": "3-10 PM", "Wednesda...
10	https://www...	SIX ON SIX CAFE - BRUNCH & RESTAURANT	("Saturday": "7 AM-10 PM", "Sunday": "3-10 PM", "Monday": "3-11 PM", "Tuesday": "3-10 PM", "Wednesda...
11	https://www...	Mỹ Quảng Cỏ Sầu	("Saturday": "6 AM-11 PM", "Sunday": "6 AM-11 PM", "Monday": "6 AM-11 PM", "Tuesday": "6 AM-11 PM", "...
12	https://www...	Thia Gỗ Restaurant Da Nang	Friday 10AM-10PM Saturday 10AM-10PM Sunday 10AM-10PM Monday 10AM-10PM Tuesday 1...
13	https://www...	EMO'S HOMECOOKED VIETNAMESE CUISINE	("Saturday": "6 AM-11 PM", "Sunday": "6 AM-11 PM", "Monday": "6 AM-11 PM", "Tuesday": "6 AM-11 PM", "...
14	https://www...	Restaurant PHỞ HOUSE	("Saturday": "10 AM-10 PM", "Sunday": "10 AM-10 PM", "Monday": "10 AM-10 PM", "Tuesday": "10 AM-10 ...
15	https://www...	Nhà Hàng Chay Ngọc Chi - Vegetarian Restaurant Mi...	("Saturday": "7 AM-9 PM", "Sunday": "7 AM-9 PM", "Monday": "7 AM-9 PM", "Tuesday": "7 AM-9 PM", "Wed...
16	https://www...	Anchor Beer Club	("Saturday": "10 AM-12 AM", "Sunday": "10 AM-12 AM", "Monday": "10 AM-12 AM", "Tuesday": "10 AM-12 A...
17	https://www...	Nhà Hàng Đà Nẵng - Ho Xanh Restaurant	("Saturday": "9 AM-10:30 PM", "Sunday": "9 AM-10:30 PM", "Monday": "9 AM-10:30 PM", "Tuesday": "9 AM...
18	https://www...	Benaras Central Da Nang	("Saturday": "11 AM-11 PM", "Sunday": "11 AM-11 PM", "Monday": "11 AM-11 PM", "Tuesday": "11 AM-11 ...

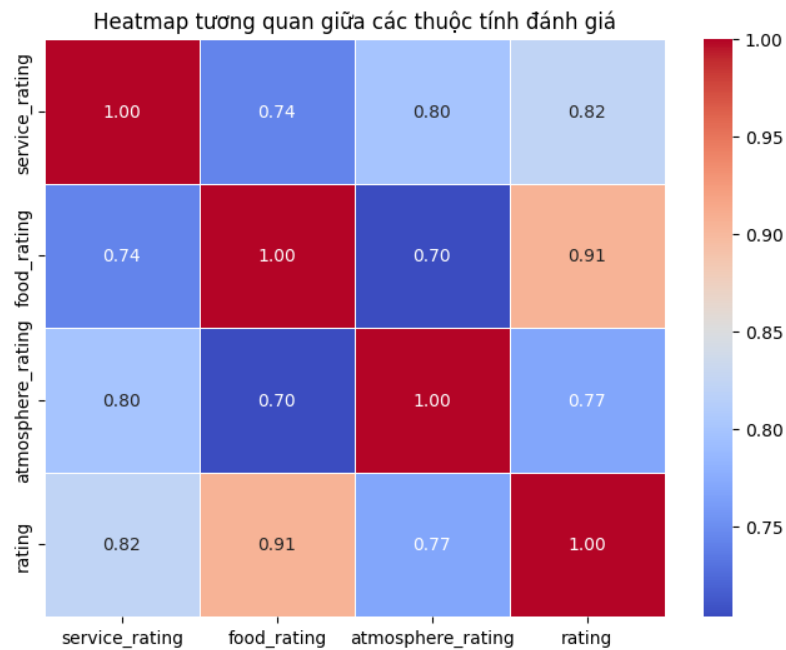
b. Xử lý giá trị bị thiếu:

- Bảng Restaurants:

- Kết quả sau khi cào có các cột sau bị thiếu dữ liệu:
 - Cột “Opening_hours” thiếu 2 dòng.
 - Cột “Phone” thiếu 34 dòng.
 - Cột “accessibility_info” thiếu 4 dòng.
- Các 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 vì vậy nhóm thay chúng bằng giá trị “Unknown” tức là không có giá trị.

- Bảng Reviews:

- Kết quả sau khi cào có các cột sau bị thiếu dữ liệu:
 - Cột “reviewer_info” thiếu 126 dòng.
 - Cột “service_rating” thiếu 2874 dòng.
 - Cột “food_rating” thiếu 2885 dòng.
 - Cột “atmosphere_rating” thiếu 2903 dòng.
 - Cột “service_type” thiếu 7322 dòng.
 - Cột “meal_type” thiếu 7271 dòng.
 - Cột “nation” thiếu 7353 dòng.
- Tương tự với bảng Restaurants, các cột reviewer_info, service_type, meal_type, nation 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 nhóm thực hiện đánh giá mức độ tương quan của chúng với cột rating.



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

2.2. Rời rạc hoá dữ liệu

Dữ liệu thời gian mở cửa khi nào về nhóm thực hiện chuyển đổi thành JSON trước khi thêm chúng vào CSDL. Kết quả khi cào:

title	rating	price_level	restaurant_type	opening_hours
Nhà hàng NHÀ BẾP CHỢ HÀN	4.7	₫100-200K	Vietnamese restaurant	Thursday 9 AM-10 PM Friday 9 AM-10 PM Saturday 9 AM-10 PM Sunday 9 AM-10 PM Monday 9...
Redcrab Danang	4.8	Unknown	Traditional restaurant	Friday 10 AM-10 PM Saturday 10 AM-10 PM Sunday 10 AM-10 PM Monday 10 AM-10 PM Tuesd...
Nhà Bếp Xúa Restaurant	4.8	₫100-200K	Vietnamese restaurant	Friday 10 AM-9:30 PM Saturday 10 AM-9:30 PM Sunday 10 AM-8 PM Monday 10 AM-9:30 PM Tu...
LoCo Restaurant - Danang seafood restaurant - 다남 ...	4.8	₫đ	Restaurant	Friday 10:30 AM-12 AM Saturday 10:30 AM-12 AM Sunday 10:30 AM-12 AM Monday 10:30 AM-12...
Nhà Hàng Si Dining	4.9	₫1,000,000+	Italian restaurant	("Friday": "5:30-9:30 PM", "Saturday": "5:30-9:30 PM", "Sunday": "5:30-9:30 PM", "Monday": "5:30-9:3...
Gypsy Rooftop Restaurant & Bar	4.9	₫1,000,000+	Restaurant	("Friday": "5:30-9:30 PM", "Saturday": "5:30-9:30 PM", "Sunday": "5:30-9:30 PM", "Monday": "5:30-9:3...
Nhà hàng Lân Gió	4.9	Unknown	Vietnamese restaurant	("Friday": "10:30 AM-10:30 PM", "Saturday": "10:30 AM-10:30 PM", "Sunday": "10:30 AM-10:30 PM", "...
Nhà hàng Thiên Kim	4.8	₫100-300K	Restaurant	("Friday": "10:30 AM-10:30 PM", "Saturday": "10:30 AM-10:30 PM", "Sunday": "10:30 AM-10:30 PM", "...
MeliOh Bistro Danang	4.6	Unknown	Restaurant	("Saturday": "7 AM-10 PM", "Sunday": "3-10 PM", "Monday": "3-11 PM", "Tuesday": "3-10 PM", "Wedn...
SIX ON SIX CAFE - BRUNCH & RESTAURANT	4.8	₫100-200K	Restaurant	("Saturday": "7 AM-10 PM", "Sunday": "3-10 PM", "Monday": "3-11 PM", "Tuesday": "3-10 PM", "Wedn...

Để chuỗi dữ liệu dài như vậy sẽ rất khó để trực quan, do đó nhóm tách chúng thành giờ mở cửa của từng ngày. Kết quả sau khi xử lý:

restaurant_id	restaurants_name	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
1	Nhà hàng NHÀ BẾP CHỢ HÀN	9 AM-10 PM	9 AM-10 PM	9 AM-10 PM	9 AM-10 PM	9 AM-10 PM	9 AM-10 PM	9 AM-10 PM
2	Thịa Gỗ Restaurant Da Nang	10 AM-10 PM	10 AM-10 PM	10 AM-10 PM	10 AM-10 PM	10 AM-10 PM	10 AM-10 PM	10 AM-10 PM
3	Bếp Cuốn Đà Nẵng	10:30 AM-9 PM	10:30 AM-9 PM	10:30 AM-9 PM	10:30 AM-9 PM	10:30 AM-9 PM	10:30 AM-9 PM	10:30 AM-9 PM
4	Có Ba Phở bò	10 AM-10 PM	10 AM-10 PM	10 AM-10 PM	10 AM-10 PM	10 AM-10 PM	10 AM-10 PM	10 AM-10 PM
5	Nhà Hàng Nhà Bếp Khuê Mỹ	10 AM-10 PM	10 AM-10 PM	10 AM-10 PM	10 AM-10 PM	10 AM-10 PM	10 AM-10 PM	10 AM-10 PM
6	Sky View Restaurant	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown
7	Ấn Thái Restaurant	10:30 AM-9:30 PM	10:30 AM-9:30 PM	10:30 AM-9:30 PM	10:30 AM-9:30 PM	10:30 AM-9:30 PM	10:30 AM-9:30 PM	10:30 AM-9:30 PM
8	Red Sky Steakhouse	11 AM-2 PM	11 AM-2 PM	11 AM-2 PM	11 AM-2 PM	11 AM-2 PM	11 AM-2 PM	11 AM-2 PM
9	MAAZI Da Nang	11 AM-12 AM	11 AM-12 AM	11 AM-12 AM	11 AM-12 AM	11 AM-12 AM	11 AM-12 AM	11 AM-12 AM
10	Redcrab Danang	10 AM-10 PM	10 AM-10 PM	10 AM-10 PM	10 AM-10 PM	10 AM-10 PM	10 AM-10 PM	10 AM-10 PM
11	Nhà Bếp Xúa Restaurant	10 AM-9:30 PM	10 AM-9:30 PM	10 AM-9:30 PM	10 AM-9:30 PM	10 AM-9:30 PM	10 AM-9:30 PM	10 AM-8 PM
12	LoCo Restaurant - Danang seafood restaurant - 다남 ...	10:30 AM-12 AM	10:30 AM-12 AM	10:30 AM-12 AM	10:30 AM-12 AM	10:30 AM-12 AM	10:30 AM-12 AM	10:30 AM-12 AM
13	Nhà Hàng Si Dining	5:30-9:30 PM	5:30-9:30 PM	5:30-9:30 PM	5:30-9:30 PM	5:30-9:30 PM	5:30-9:30 PM	5:30-9:30 PM

Dữ liệu cột “accessibility_info” khi cào về được định dạng thành dữ liệu JSON trước khi thêm vào CSDL trên SQL Server. Kết quả trước khi xử lý:

id	tên	accessibility_info
1	Nhà hàng NHÀ BẾP CHỢ HẢI	[{"No wheelchair-accessible car park", "No wheelchair-accessible entrance", "No wheelchair-accessible seating", "Offers takeaway", "Serves dine-in", "Has great tea selection", "Popular for breakfast", "Popular for lunch", "Popular for dinner", "Good for solo dining", "Good for groups", "Has outdoor seating", "Has no-contact delivery", "Offers delivery", "Offers take..."}]
2	Thia Gõ Restaurant Da Nang	[{"Has wheelchair-accessible seating", "Does not have assistive hearing loop", "No wheelchair-accessible car park", "No wheelchair-accessible entrance", "Has outdoor seating", "Has no-contact delivery", "On-site services available", "Offers takeaway", "Serves dine-in", "N..."}]
3	Bếp Cuốn Đà Nẵng	[{"Has wheelchair-accessible seating", "Has wheelchair-accessible toilet", "No wheelchair-accessible car park", "Has outdoor seating", "Has no-contact delivery", "Offers delivery", "Offers takeaway", "Serves dine-in", "Popular for breakfast", "Popular for lunch", "Pop..."}]
4	Cố Ba Phở bò	[{"No wheelchair-accessible car park", "No wheelchair-accessible entrance", "Has no-contact delivery", "Offers delivery", "On-site services available", "Offers takeaway", "Serves dine-in", "Popular for breakfast", "Popular for lunch", "Pop..."}]
5	Nhà Hàng Nhà Bếp Khuê Mỹ	[{"No wheelchair-accessible car park", "No wheelchair-accessible entrance", "No wheelchair-accessible seating", "Identifies as women-owned", "Has no-contact delivery", "Offers delivery", "Offers takeaway", "Serves dine-in", "Pop..."}]
6	Sky View Restaurant	[{"Has wheelchair-accessible entrance", "Has wheelchair-accessible toilet", "Has outdoor seating", "Offers takeaway", "Serves dine-in", "Serves great cocktails", "Has rooftop seating", "Popular for dinner", "Serves alcohol", "Serves..."}]
7	Ăn Thôi Restaurant	[{"Has wheelchair-accessible toilet", "No wheelchair-accessible car park", "No wheelchair-accessible seating", "Has no-contact delivery", "Offers delivery", "Offers takeaway", "Serves dine-in", "Popular for breakfast", "Popular for lu..."}]
8	Red Sky Steakhouse	[{"No wheelchair-accessible car park", "Has no-contact delivery", "Offers delivery", "On-site services available", "Offers takeaway", "Serves dine-in", "Serves great cocktails", "Popular for lunch", "Popular for dinner", "Good for solo di..."}]
9	MAAZI Da Nang	[{"Has wheelchair-accessible seating", "Has wheelchair-accessible toilet", "Offers kerbside pickup", "Has no-contact delivery", "Offers delivery", "Offers takeaway", "Serves dine-in", "Has great tea selection", "Popular for lunch", "Po..."}]
10	Redonab Danang	[{"Has wheelchair-accessible car park", "Has wheelchair-accessible entrance", "Has wheelchair-accessible seating", "Has wheelchair-accessible toilet", "Does not have assistive hearing loop", "Offers kerbside pickup", "Has no-c..."}]
11	Nhà Bếp Xưa Restaurant	[{"Has wheelchair-accessible seating", "No wheelchair-accessible car park", "Has outdoor seating", "Offers kerbside pickup", "Has no-contact delivery", "Offers delivery", "Offers takeaway", "Serves dine-in", "Has great tea selection..."}]
12	LoCo Restaurant - Danang seafood restaurant - Chợ	[{"Has wheelchair-accessible toilet", "Has outdoor seating", "Offers kerbside pickup", "Has no-contact delivery", "Offers delivery", "Has drive-through", "Offers takeaway", "Serves dine-in", "Serves great cocktails", "Has live music", "H..."}]

Việc để dữ liệu thành chuỗi dài như vậy sẽ rất khó để thực hiện trực quan hoá, cũng như phân tích huấn luyện mô hình sau này, do đó nhóm quyết định tách chúng ra thành bảng nhỏ hơn và được liên kết với bảng “Restaurants” thông qua khoá chính là “restaurant_id”.

Vì mối quan hệ giữa nhà hàng và tiện ích là mối quan hệ nhiều – nhiều, do đó khi thực hiện trên CSDL nhóm tạo thêm bảng trung gian là “features_item”. Kết quả sau khi xử lý:

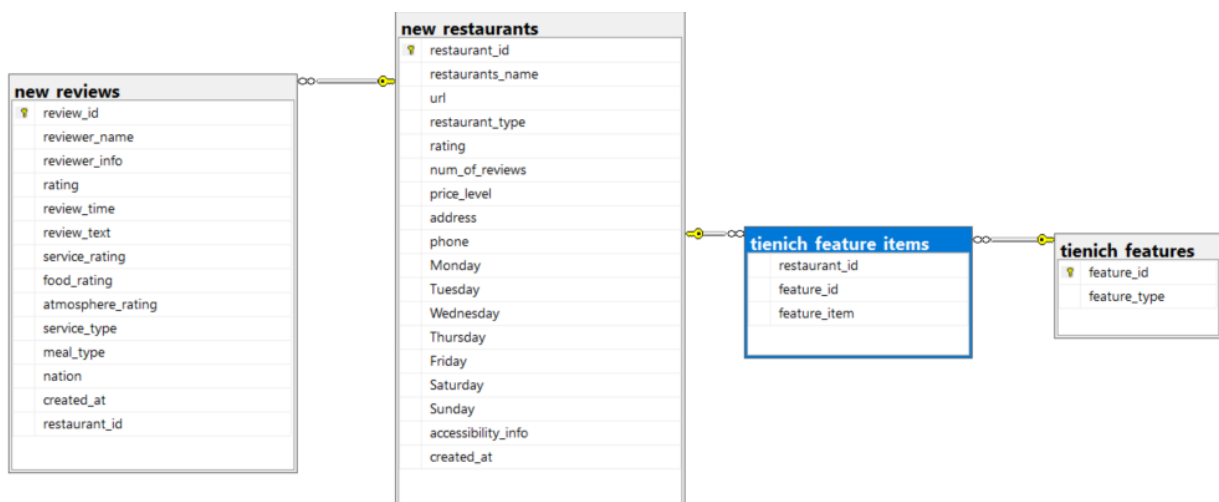
- **Bảng Feature:** chứa tên các loại tiện ích của nhà hàng bao gồm 11 loại

	feature_id	feature_type
1	1	accessibility_info
2	2	service_options
3	3	offerings
4	4	dining_options
5	5	amenities
6	6	atmosphere
7	7	crowd
8	8	popular_for
9	9	payments
10	10	children
11	11	parking

- **Bảng Feature_item:** chứa tên các tiện ích cụ thể của từng loại tiện ích, có khoá ngoại được liên kết với “restaurant_id” và “feature_id”

restaurant_id	feature_id	feature_item
1	1	No wheelchair-accessible car park
1	1	No wheelchair-accessible entrance
1	1	No wheelchair-accessible seating
1	2	Offers takeaway
1	2	Serves dine-in
1	3	Quick bite
1	3	Serves alcohol
1	3	Serves beer
1	3	Serves coffee
1	3	Serves small plates
1	4	Has catering
1	4	Has seating
1	4	Serves breakfast
1	4	Serves brunch
1	4	Serves dessert
1	4	Serves dinner
1	4	Serves lunch
1	5	Free Wi-Fi

- Sơ đồ quan hệ - thực thể trong CSDL:



2.3. Biến đổi dữ liệu

a. Bảng Reviews

Cột “review_time” khi cào về chỉ chứa thông tin thời gian từ ngày thực khách viết đánh giá cho đến thời điểm thu thập dữ liệu, thông tin này không có ý nghĩa khi trực quan và huấn luyện mô hình. Vì thế nhóm sử dụng Python để tính ngày thực khách viết đánh giá dựa trên giá trị của cột thời gian thu thập dữ liệu lùi mốc thời gian khớp với thông tin của cột “review_time”. Kết quả trước khi xử lý:

	review_id	reviewer_name	rating	review_time	review_text
1	RV00001	Joe Greco	5	3 weeks ago	As soon as the food arrived I knew I'd made the right ...
2	RV00002	Annie Song	5	3 weeks ago	As Vietnamese living overseas, we loved the flavours ...
3	RV00003	Jiwon Kim	5	2 months ago	It was a huge and cozy place to try Vietnam style food!...
4	RV00004	katsie tang	3	5 months ago	Had the beef pho and fried rice. Flavors not too authe...
5	RV00005	Riina Runnel	5	5 months ago	I would like to give more stars! I'm a big fan of this pla...
6	RV00006	Richard Hinchman	3	2 months ago	Very Koreanized version of Vietnamese food. We cam...
7	RV00007	Elk T	4	3 months ago	Tourist-friendly place to eat. Clean and accommodati...
8	RV00008	Emily Lee	4	3 months ago	Service is good. Staff are friendly and attentive. The fo...
9	RV00009	Erng Mitphumwiboon	5	2 years ago	Foods are nice and the portion is generous. We order...
10	RV00010	ThanhHa H.	3	a month ago	Located across from Chợ Hàn... very busy at lunch tim...

Kết quả sau khi đã xử lý với Python:

	review_id	reviewer_name	rating	review_date	review_text
0	RV00001	Joe Greco	5	2025-02-27	As soon as the food arrived I knew I'd made th...
1	RV00002	Annie Song	5	2025-02-27	As Vietnamese living overseas, we loved the fl...
2	RV00003	Jiwon Kim	5	2025-01-19	It was a huge and cozy place to try Vietnam st...
3	RV00004	katsie tang	3	2024-10-21	Had the beef pho and fried rice.\r\nFlavors no...
4	RV00005	Riina Runnel	5	2024-10-22	I would like to give more stars! I'm a big fan...

b. Bảng Restaurants

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	restaurants_name	address	district
1	Nhà hàng NHÀ BẾP CHỢ HÀN	22 Hùng Vương, Hải Châu 1, Hải Châu, Đà Nẵng 5...	Hải Châu
2	Thìa Gỗ Restaurant Da Nang	53 Phan Thúc Duyệt, Bắc Mỹ Phú, Ngũ Hành Sơn, ...	Ngũ Hành Sơn
3	Bếp Cuốn Đà Nẵng	31-33 Trần Bạch Đằng, Phước Mỹ, Sơn Trà, Đà Nẵ...	Sơn Trà
4	Cô Ba Phở bò	154 Bạch Đằng, Hải Châu 1, Hải Châu, Đà Nẵng 5...	Hải Châu
5	Nhà Hàng Nhà Bếp Khuê Mỹ	416 Võ Nguyên Giáp, Bắc Mỹ An, Ngũ Hành Sơn, Đ...	Ngũ Hành Sơn
6	Sky View Restaurant	216 Võ Nguyên Giáp, Phước Mỹ, Sơn Trà, Đà Nẵng...	Sơn Trà
7	Ăn Thôi Restaurant	114 Bạch Đằng, Hải Châu 1, Hải Châu, Đà Nẵng 5...	Hải Châu
8	Red Sky Steakhouse	248 Trần Phú, Phước Ninh, Hải Châu, Đà Nẵng 55...	Hải Châu
9	MAAZI Da Nang	264 Trần Phú, Phước Ninh, Hải Châu, Đà Nẵng 70...	Hải Châu
10	Redcrab Danang	138 Phạm Văn Đồng, street, Sơn Trà, Đà Nẵng 55...	Sơn Trà

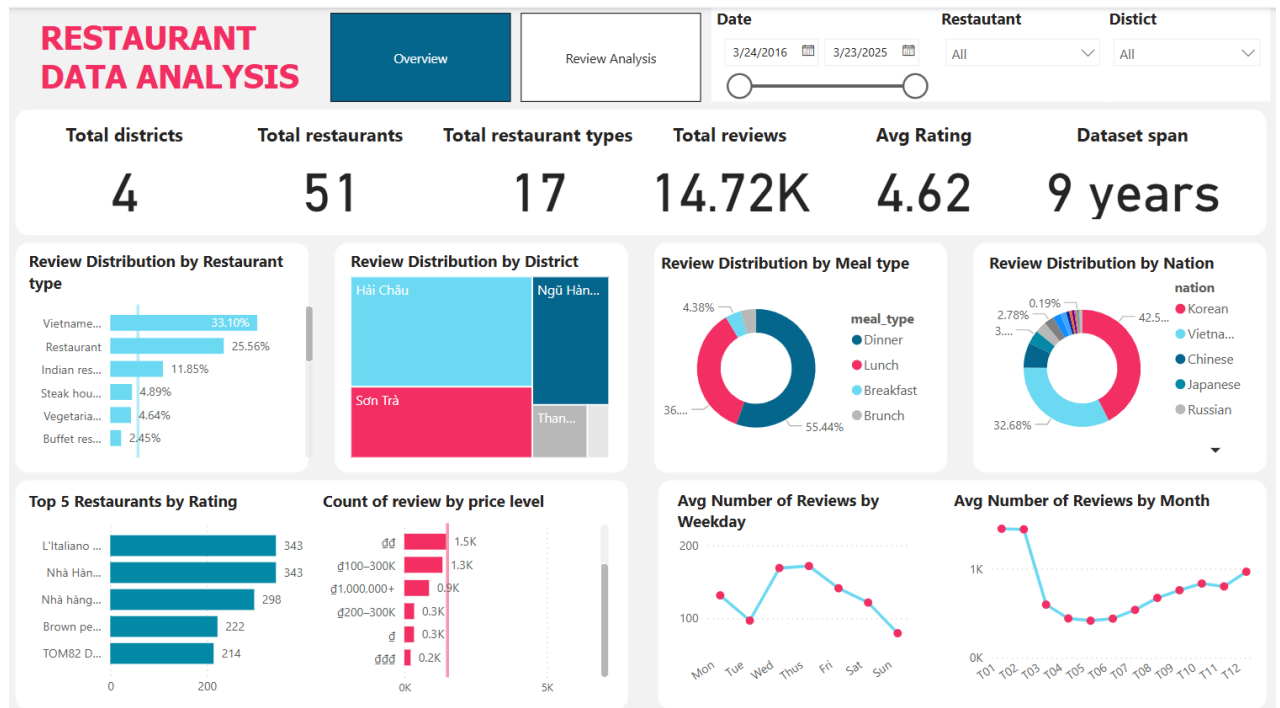
c. Gắn nhãn reviews

PHẦN 4: TRỰC QUAN HOÁ DỮ LIỆU

1. Sơ đồ quan hệ thực thể

Biểu đồ phân tích này cung cấp thông tin chi tiết từ 14.720 đánh giá về 51 nhà hàng tại Đà Nẵng, trải dài trong 9 năm (từ 2016 đến 2025). Đây là những thông tin hữu ích cho các startup muốn mở nhà hàng mới hoặc các chủ nhà hàng hiện tại muốn cải thiện hiệu quả kinh doanh.

1. Trang Overview



Trang Overview cung cấp cái nhìn tổng thể về bối cảnh nhà hàng tại Đà Nẵng, giúp các bên liên quan (các chủ nhà hàng, các bên chuẩn bị mở nhà hàng đang tìm kiếm chiến lược phát triển nhà hàng phù hợp) hiểu được xu hướng chung, mức độ cạnh tranh, và đặc điểm của thị trường ẩm thực dựa trên dữ liệu nhà hàng.

1. Trang Review Analysis

