Internship Project Plan – Restaurant Recommendation System

1. `Users`

Column Name	Type	Description
`user_id`	INT (PK)	Unique user ID
`name`	STRING	Full name
`birth_year`	INT	Year of birth
`gender`	STRING	'Male' or 'Female'
`latitude`	FLOAT	User's location (latitude)
`longitude`	FLOAT	User's location (longitude)
`city`	STRING	City name

---

2. `MedicalCondition`

Column Name	Type	Description
`condition_id`	INT (PK)	Unique ID for each condition
`name`	STRING	e.g., Diabetes, Hypertension, etc.

---

3. `UserMedicalCondition` (Mapping)

Column Name	Type	Description
`user_id`	INT (FK)	From `Users`
`condition_id`	INT (FK)	From `MedicalCondition`

---

4. `Restaurants`

Column Name	Type	Description
`restaurant_id`	INT (PK)	Unique restaurant ID
`name`	STRING	Name of the restaurant
`pet_friendly`	BOOLEAN	Whether it allows pets

`latitude`	FLOAT	Latitude of location
`longitude`	FLOAT	Longitude of location
`city`	STRING	City name
`rating`	FLOAT	Aggregate rating

---

5. `RestaurantMenu`

Combines `Cuisine`, `FoodPreference`, and `FoodEaten`.

Column Name	Type	Description
`menu_id`	INT (PK)	Unique ID for each menu item
`restaurant_id`	INT (FK)	From `Restaurants`
`dish_name`	STRING	Name of the dish
`cuisine`	STRING	e.g., Indian, Italian
`is_vegetarian`	BOOLEAN	Dish is vegetarian
`is_vegan`	BOOLEAN	Dish is vegan

---

6. `VisitHistory`

Column Name	Type	Description
`visit_id`	INT (PK)	Unique visit ID
`user_id`	INT (FK)	Who visited
`restaurant_id`	INT (FK)	Which restaurant

rating_given	FLOAT	Rating for that visit (1.0 to 5.0)
visit_date	DATE	Date of the visit
feedback	TEXT	Optional feedback (~30% rows)

---

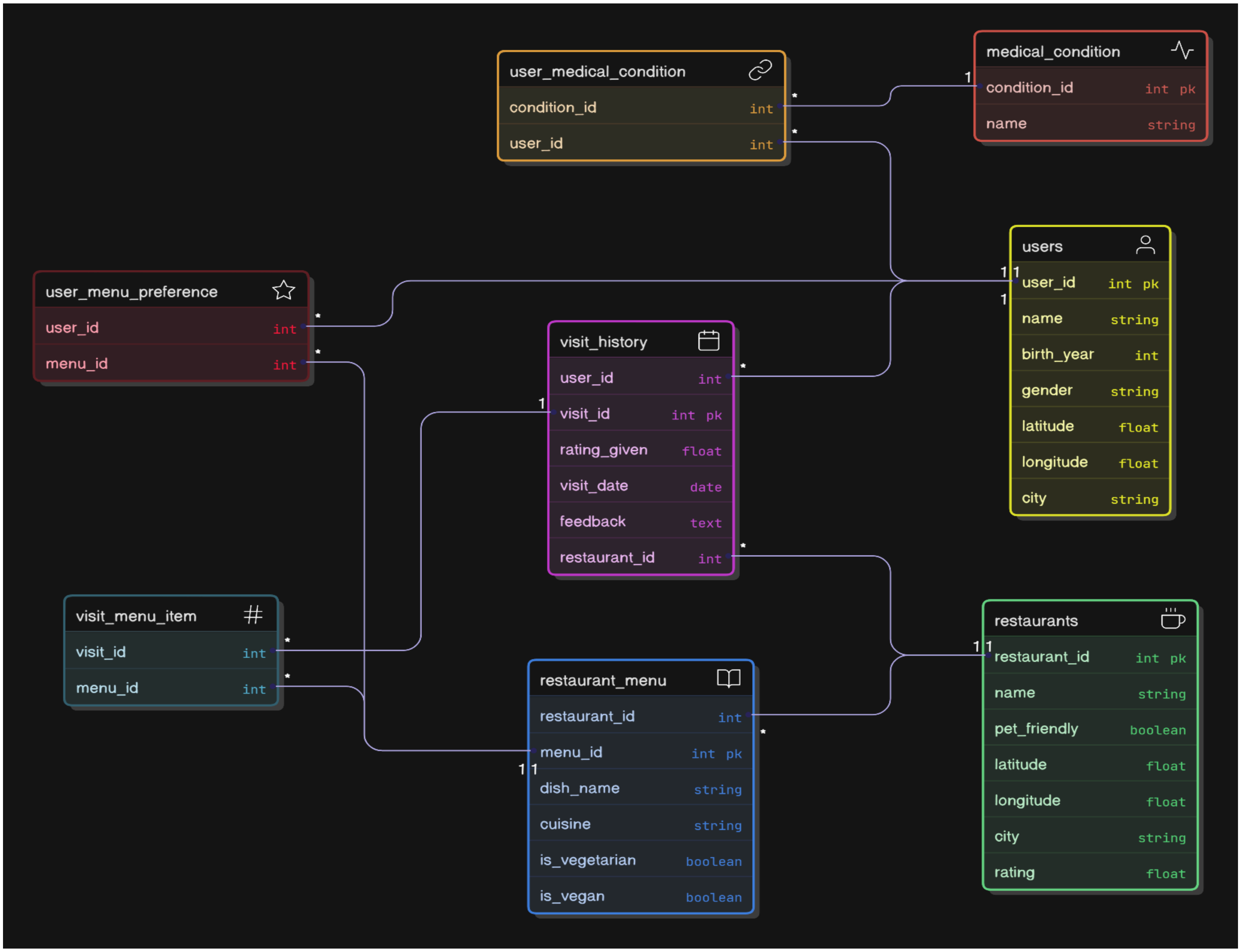
7. VisitMenuItem (Mapping)

Maps what food was eaten during each visit

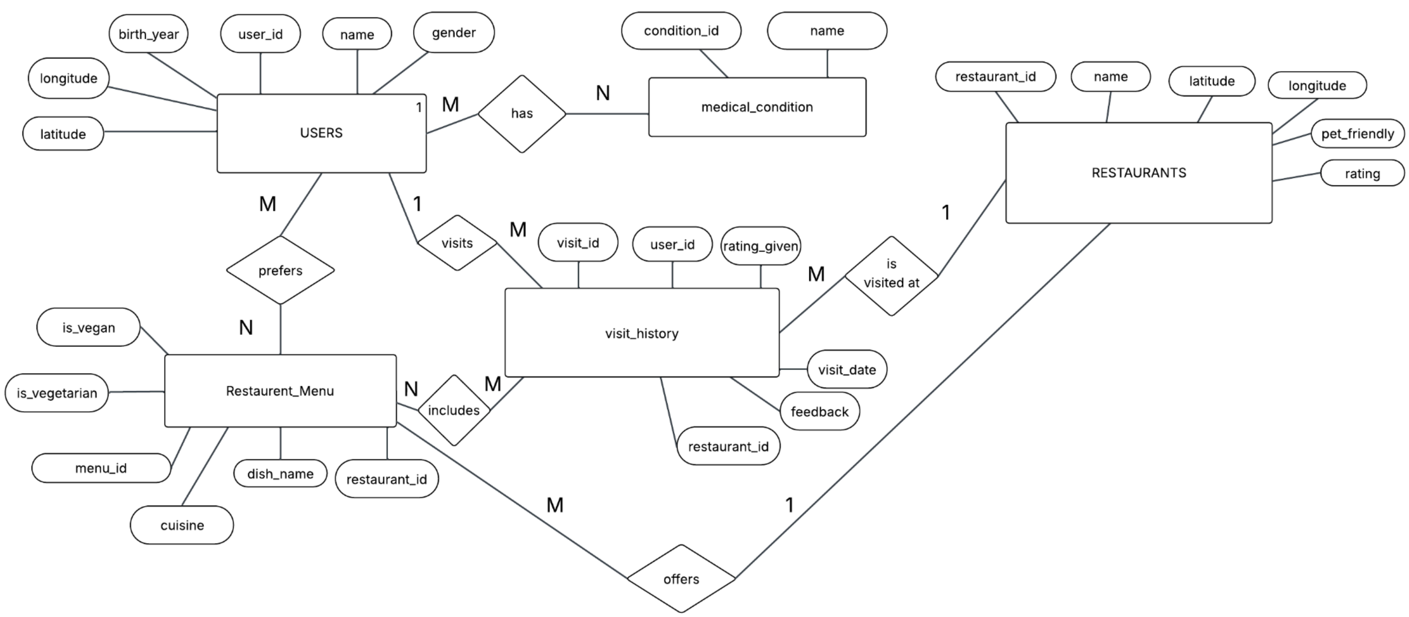
Column Name	Type	Description
visit_id	INT (FK)	From VisitHistory
menu_id	INT (FK)	From RestaurantMenu

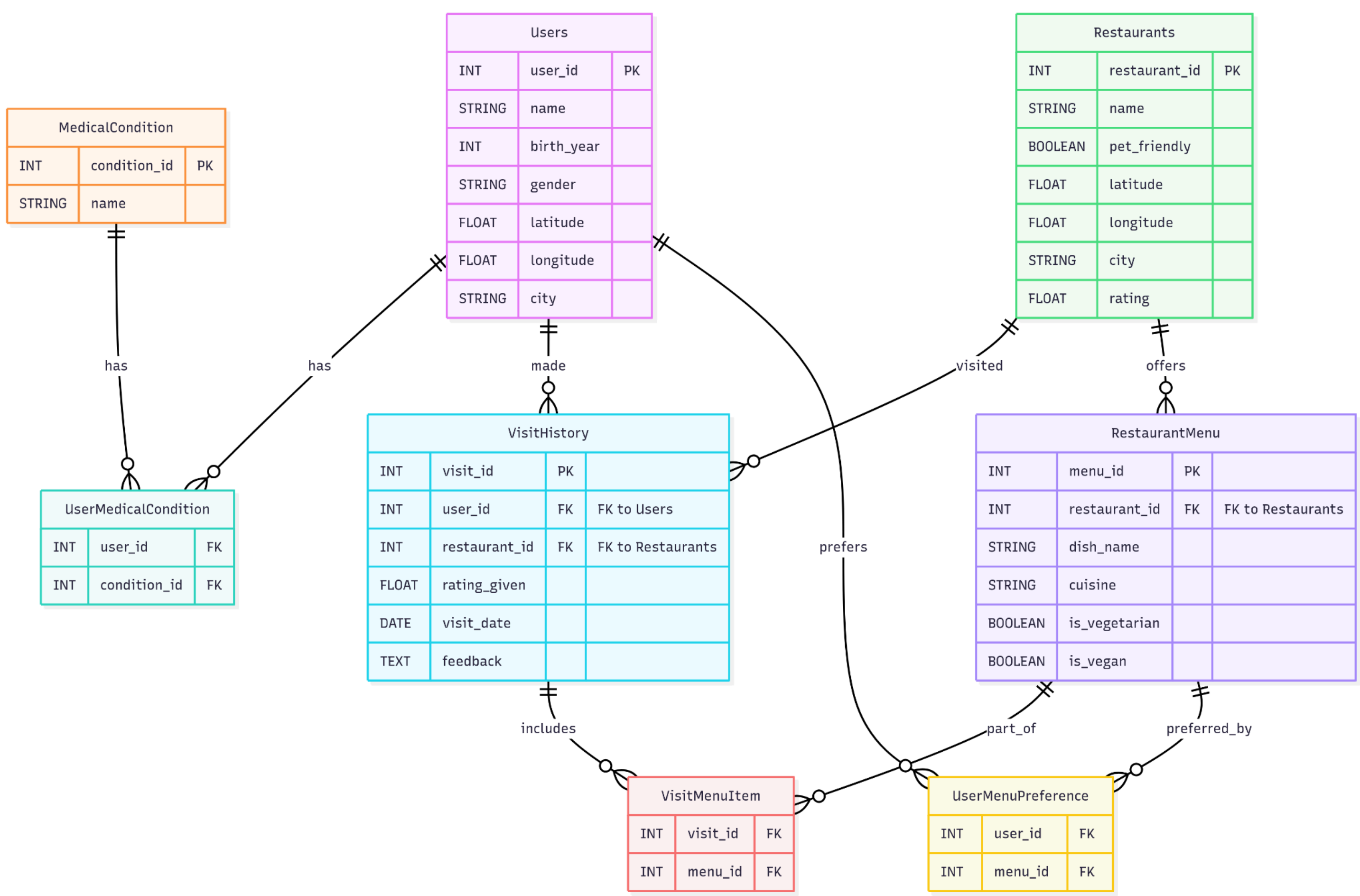
UserMenuPreference

Column Name	Type	Description
user_id	INT (FK)	From Users
menu_id	INT (FK)	From RestaurantMenu (dish/item they like)


user_medical_condition
condition_id int
user_id int
medical_condition
condition_id int pk
name string
users
user_id int pk
name string
birth_year int
gender string
latitude float
longitude float
city string
user_menu_preference
user_id int
menu_id int
visit_history
user_id int
visit_id int pk
rating_given float
visit_date date
feedback text
restaurant_id int
visit_menu_item
visit_id int
menu_id int
restaurant_menu
restaurant_id int
menu_id int pk
dish_name string
cuisine string
is_vegetarian boolean
is_vegan boolean
restaurants
restaurant_id int pk
name string
pet_friendly boolean
latitude float
longitude float
city string
rating float

Laser pointer Share





2. Tools and Libraries Used

- **Python**: Core language for data orchestration
- **Faker**: Generates realistic names, locations, companies
- **Pandas**: Data structure management and DataFrame operations
- **NumPy**: Controlled randomness and numerical calculations
- **Datetime**: Time generation for historical visit data
- **SQLAlchemy**: ORM for SQL Server integration
- **Google Gemini API**: Generates natural language dish names and reviews

3. Data Tables and Schema Overview

1. Users Table

- **Rows**: 100,000
- **Key Features**:
  - Balanced gender ratio
  - Age range 18–75 (birth year 1950–2007)
  - Location within 15 Indian cities
  - Coordinates jittered within ±0.05° of city center

---

2. UserMedicalCondition Table

- **Rows**: ~120,000
- **Key Features**:
  - Each user has 0–2 conditions
  - “None” condition (id=1) added only if no real condition

---

3. MedicalCondition Table

- **Rows**: 20 (customizable)
- **Includes**: 'None', 'Diabetes', etc.

---

4. Restaurants Table

- **Rows**: 1,000
- **Key Features**:
  - Random names with realistic suffixes (Bistro, Cafe)
  - Ratings between 2.0 and 5.0
  - 30% are pet-friendly

---

5. RestaurantMenu Table

- **Rows**: 500 (adjustable limit)
- **Key Features**:
  - Dish names generated via Gemini based on cuisine
  - Veg/vegan flags based on dish name ingredients

---

6. VisitHistory Table

- **Rows**: 100,000
- **Key Features**:
  - Random visit dates over 5 years

    - Feedback in ~30% of rows
    - Feedback sentiment depends on rating

---

7. VisitMenuItem Table

- **Rows**: ~300,000+
- **Key Features**:
    - Each visit includes 1–5 menu items (no repeats)

---

8. UserMenuPreference Table

- **Rows**: ~200,000
- **Key Features**:
    - Each user has 1–10 preferred menu items

---

4. Key Features & Constraints

- **Geo Accuracy**: ±0.05° variation from real coordinates
- **Feedback Logic**:
    - 5.0 → positive
    - 3.0 → neutral
    - <2.5 → negative
- **Medical Condition Logic**:
    - “None” only when no actual condition
- **Menu Sampling**: No duplicate dishes per visit

- **Normalization**:
  - Separate tables for mappings
  - No redundant fields
- **Scalability**:
  - Designed for 100k+ users and visits
  - Handles referential integrity across all tables

5. Observed Edge Cases & Solutions

Edge Case	Solution
Unrealistic feedback	Used sentiment-aligned Gemini prompts
Overrepresentation of feedback	Restricted to ~30% of visit entries
Disjointed coordinates	Jittered around exact city lat/lon from a fixed list
'None' co-existing with other conditions	Ensured 'None' only added when no other condition exists
Overload of Gemini prompts for menus/feedback	Capped menu items and used prefilled feedback pool

6. File Outputs

- All tables saved to **SQL Server** using SQLAlchemy engine
- Structured and documented schema with normalization
- Data can be exported for analytics, training, or ML

RestaurentMenu table:

100 %

Results | Messages

	menu_id	restaurant_id	dish_name	cuisine	is_vegetarian	is_vegan
1	1	1	Tagine with Apricots and Lamb	Moroccan	0	0
2	2	1	Doro Wot	Ethiopian	1	1
3	3	1	Carciofi alla Romana	Italian	1	1
4	4	1	Moqueca de Camarão com Dendê	Brazilian	1	1
5	5	1	Samkeh Harra	Lebanese	1	1
6	6	1	Spicy Muhammara with Pomegranate and Walnuts	Lebanese	1	1
7	7	1	Melitzanosalata	Greek	1	1
8	8	1	Tagine with Apricots and Lamb	Moroccan	0	0
9	9	1	Kinoko to Nasu no Nimono	Japanese	1	1
10	10	1	Gambas al Ajillo	Spanish	1	1
11	11	1	Salmone al Limone e Rosmarino	Italian	1	1
12	12	1	Schweinshaxe	German	1	1
13	13	2	Sole Meunière	French	1	1
14	14	2	Saltimbocca alla Romana	Italian	1	1
15	15	2	Tinga Poblana	Mexican	1	1
16	16	2	Labneh with Za'atar	Lebanese	1	1

Users Table:

100 %

Results | Messages

	user_id	name	birth_year	gender	latitude	longitude	city
1	1	David Merritt	1953	Male	12.279425	76.643157	Mysore
2	2	John Jimenez	1992	Female	12.921973	74.855592	Mangalore
3	3	David Nicholson	1963	Female	13.353254	74.741214	Udupi
4	4	Michael Andrews	1961	Male	31.586894	74.893002	Amritsar
5	5	Joseph Coleman	1982	Female	13.316222	74.703567	Udupi
6	6	Maria Petersen	2005	Female	9.935476	76.276394	Kochi
7	7	Amanda Anderson	1986	Female	9.963708	76.226926	Kochi
8	8	Anthony Jones	1977	Male	13.072239	80.291888	Chennai
9	9	Joshua Owen	1960	Male	11.250497	75.780336	Kozhikode
10	10	Jeremy Adams	1961	Female	31.593655	74.867186	Amritsar
11	11	Ricky Collier	1967	Male	13.30452	74.712136	Udupi
12	12	Mike Flores	1981	Male	13.315127	74.766126	Udupi
13	13	Darrell Bryant	1979	Male	31.609765	74.883525	Amritsar
14	14	Paul Cantrell	1986	Female	19.107044	72.872661	Mumbai
15	15	Spencer Thompson	1982	Female	8.56254	76.905787	Trivandrum
16	16	Dr. Jessica Campbell	1967	Female	17.388392	78.529625	Hyderabad

Restaurant Table:

100 %

Results | Messages

	restaurant_id	name	pet_friendly	latitude	longitude	city	rating
1	1	Carter Ltd Restaurant	1	22.611825	88.384036	Kolkata	4.2
2	2	Simmons-Fitzpatrick Restaurant	0	13.081392	80.228343	Chennai	3.3
3	3	Larson, Walker and Buchanan Restaurant	0	12.902394	74.813637	Mangalore	2.4
4	4	Atkins-Santiago Restaurant	0	12.98579	77.607778	Bangalore	2.5
5	5	Garza, Adams and Lewis Restaurant	0	22.537968	88.354827	Kolkata	4.8
6	6	Nguyen-Stewart Grill	1	12.3417	76.644186	Mysore	4.2
7	7	Neal, Williams and Murray Cafe	0	9.924202	76.311995	Kochi	2.4
8	8	Elliott Ltd Cafe	0	8.505992	76.95722	Trivandrum	2.7
9	9	Carlson, Livingston and Santos Cafe	0	31.671945	74.882606	Amritsar	3.8
10	10	Soto-Burns Bistro	0	17.390105	78.474605	Hyderabad	3.7
11	11	Todd, Hunter and Pitts Cafe	0	8.550519	76.964677	Trivandrum	3.1
12	12	Barnes-Huang Grill	0	19.06892	72.860337	Mumbai	2.7
13	13	Williams Group Grill	0	18.472557	73.89814	Pune	4.8
14	14	Massey-Hill Bistro	1	12.906287	74.819058	Mangalore	3.9
15	15	Simmons-Wagner Restaurant	0	13.386869	74.731941	Udupi	3.9
16	16	Webb-Nguyen Cafe	0	31.665815	74.91106	Amritsar	2

Visit History table:

100 %

Results | Messages

	visit_id	user_id	restaurant_id	rating_given	visit_date	feedback
37	37	20648	140	3.2	2025-04-08	The restaurant served standard fare in a clean setting.
38	38	12053	239	4.9	2022-06-08	NULL
39	39	82536	294	3.1	2020-08-12	NULL
40	40	49483	528	2.7	2022-01-20	NULL
41	41	1822	182	4.9	2020-11-30	NULL
42	42	7968	972	2.4	2024-06-02	NULL
43	43	4190	824	4.3	2021-03-14	NULL
44	44	57324	278	2.3	2024-01-05	NULL
45	45	9600	995	2.6	2023-02-22	The restaurant served standard fare in a clean setting.
46	46	96440	664	3.5	2020-08-25	The restaurant served standard fare in a clean and adequately maintained setting.
47	47	74560	667	3.2	2024-01-02	NULL
48	48	587	341	1.2	2022-10-21	NULL
49	49	64890	94	4.8	2021-06-06	NULL
50	50	8775	88	1.2	2021-09-23	The only thing this restaurant seasoned well was the bill.
51	51	83044	461	3.1	2020-11-13	NULL
52	52	81104	62	1.1	2023-12-03	NULL

Medical Condition Table:

	condition_id	name
1	1	None
2	2	Diabetes
3	3	Hypertension (High Blood Pressure)
4	4	Heart Disease
5	5	High Cholesterol
6	6	Celiac Disease
7	7	Crohn's Disease
8	8	Ulcerative Colitis
9	9	Irritable Bowel Syndrome (IBS)
10	10	Gastroesophageal Reflux Disease (GERD)
11	11	Kidney Disease
12	12	Liver Disease
13	13	Gallstones
14	14	Food Allergies
15	15	Lactose Intolerance
16	16	Gluten Sensitivity (Non-Celiac Gluten Sensitivity)

VisitMenuItem table:

	visit_id	menu_id
1	33	326
2	108	295
3	108	309
4	141	79
5	160	238
6	160	221
7	178	454
8	178	453
9	178	452
10	249	17
11	249	25

UserMenuPreference Table:

100 %

Results Message

	user_id	menu_id
1	19652	450
2	19652	375
3	19652	328
4	19652	453
5	19652	288
6	19652	34
7	19652	301
8	82206	289
9	82206	204
10	82206	411
11	82206	388
12	82206	97
13	82206	12
14	82206	401
15	82206	319
16	24964	184

UserMedicalCondition Table:

100 %

Results Messages

	user_id	condition_id
1	1	1
2	2	1
3	3	1
4	4	15
5	5	12
6	5	20
7	6	7
8	7	1
9	8	1
10	9	16
11	10	1
12	11	1
13	12	15
14	13	1
15	14	1
16	15	3

7. Conclusion

The generated datasets simulate a small-scale, location-bound restaurant recommendation environment. They are suitable for training or testing machine learning models, data visualization, clustering, collaborative filtering, and geo-recommendation systems.