ADVANCED QUERIES

Q1) Chatbot: “Recommend doctors for a symptom list in a city, ranked by matches”

Use-case: user describes symptoms; we infer likely specialties and rank doctors in their city by symptom overlap.

```
WITH
input_symptoms AS (
  SELECT s.symptom_id
  FROM symptom s
  WHERE s.symptom_name IN ('chest_pain','palpitations','nausea')
),
-- only diseases that have at least one of the input symptoms
relevant_diseases AS (
  SELECT DISTINCT dis.disease_id, dis.specialty_id
  FROM disease dis
  JOIN disease_symptom ds  ON ds.disease_id = dis.disease_id
  JOIN input_symptoms ins  ON ins.symptom_id = ds.symptom_id
),
-- only doctors in the target city
ny_doctors AS (
  SELECT doctor_id, first_name, last_name, specialty_id, city, state,
phone_number
  FROM doctor
  WHERE city = 'New York'
)
SELECT
  d.doctor_id,
  CONCAT_WS(' ', d.first_name, d.last_name) AS doctor_name,
  sp.specialty_name,
  d.city, d.state, d.phone_number,
  COUNT(DISTINCT ds.symptom_id) AS matched_symptoms
FROM ny_doctors d
JOIN specialty sp          ON sp.specialty_id = d.specialty_id
JOIN relevant_diseases rd  ON rd.specialty_id = d.specialty_id
JOIN disease_symptom ds    ON ds.disease_id = rd.disease_id
JOIN input_symptoms ins    ON ins.symptom_id = ds.symptom_id
GROUP BY d.doctor_id
HAVING COUNT(DISTINCT ds.symptom_id) > 0
ORDER BY matched_symptoms DESC, doctor_name
LIMIT 15;
```

Output:

doctor_id	doctor_name	specialty_name	city	state	phone_number	matched_symptoms
c81bee8f-5eb3-44f8-8f89-c675dec803a3	A. LAWRENCE ATTIA	GASTROENTEROLOGY	NEW YORK	NY	2123077210	3
896e93e5-cc92-489c-994b-16a7c42eb7cc	AADIL RAHMAN	PEDIATRIC MEDICINE	NEW YORK	NY	2123059876	3
e1946020-661a-4a79-a262-8be7e6bcbf8c	AARON ARREDONDO	EMERGENCY MEDICINE	NEW YORK	NY	2124202000	3
da866f0e-d430-4bbe-9392-cedda5e5249e	AARON ETRA	MEDICAL ONCOLOGY	NEW YORK	NY	2128448288	3
6cd3bb2c-07c3-4ef1-982b-09b30c62a1dd	AARON GREENBERG	CARDIOVASCULAR DISEASE (CARDIOLOGY)	NEW YORK	NY	2127723111	3
ae1664c2-d0f8-40ca-abf1-006d25407ac8	AARON HULTGREN	EMERGENCY MEDICINE	NEW YORK	NY	2122639700	3
8ac2110f-92a6-4ad2-ac56-67c9ad848b14	AARON LIM	EMERGENCY MEDICINE	NEW YORK	NY	2122639700	3
dd9010bc-21c3-4946-8f8e-0710d05d532d	AARON MITCHELL	MEDICAL ONCOLOGY	NEW YORK	NY	2126392000	3
a2284f62-cf65-4406-90b8-065384d370c6	AARON SAVEDOFF	PSYCHIATRY	NEW YORK	NY	2128248150	3
e9577ce1-e79e-48b0-9a0f-8dc4620a8ca1	AARON STONESTROM	MEDICAL ONCOLOGY	NEW YORK	NY	2126392000	3
a94bf42b-d60c-42e3-a90e-ff504ac66447	AARTI BHARDWAJ	MEDICAL ONCOLOGY	NEW YORK	NY	2128635599	3
bd163652-316c-48ac-bad0-d0834d63074e	AASMA SHAUKAT	GASTROENTEROLOGY	NEW YORK	NY	2122637300	3
bfcc1e2f-1b4a-43fb-b0b7-4935fb062323	ABBY MULKEEN	PSYCHIATRY	NEW YORK	NY	2122637300	3
2856c449-cb07-4bc9-850b-4619dda1972e	ABDALLAH BEANO	GASTROENTEROLOGY	NEW YORK	NY	NULL	3
7abf4fb9-1dde-462d-a687-1093d8c57233	ABDUL-AZIZ AHMED	EMERGENCY MEDICINE	NEW YORK	NY	6469627300	3

Took about 579 seconds to run (before indexing).

Q2) Ranking: “How intense is each doctor’s coverage in a city?”

Use-case: ranks doctors by the breadth of symptoms/diseases their specialty touches in a city (great for recommendation/ranking logic).

```
WITH ny_doctors AS (
  SELECT doctor_id, first_name, last_name, specialty_id, city, state,
phone_number
  FROM doctor
  WHERE city = 'New York'            -- adjust to tune runtime
    AND specialty_id IS NOT NULL
),
ny_specs AS (                         -- only the specialties we
actually need
  SELECT DISTINCT specialty_id
  FROM ny_doctors
),
spec_pairs AS (                       -- do heavy work ONCE per
specialty
  SELECT
    dis.specialty_id,
    COUNT(DISTINCT ds.disease_id, ds.symptom_id) AS symptom_pairs,
    COUNT(DISTINCT ds.disease_id)               AS diseases_covered
  FROM disease dis
  JOIN disease_symptom ds ON ds.disease_id = dis.disease_id
  JOIN ny_specs ns        ON ns.specialty_id = dis.specialty_id
  GROUP BY dis.specialty_id
)
SELECT
  d.doctor_id,
  CONCAT_WS(' ', d.first_name, d.last_name) AS doctor_name,
  d.city, d.state, d.phone_number,
  sp.symptom_pairs,
  sp.diseases_covered
FROM ny_doctors d
JOIN spec_pairs sp ON sp.specialty_id = d.specialty_id
ORDER BY sp.symptom_pairs DESC, sp.diseases_covered DESC, doctor_name
LIMIT 15;
```

Output:

	doctor_id	doctor_name	city	state	phone_number	symptom_pairs	diseases_covered
▶	1f1fbba1-9bf6-44b0-a3dc-407d8b15b4b7	ALAN COHEN	NEW YORK	NY	2126598551	895	99
	59153403-97eb-4c2a-8331-636e207aff68	ALAN POLLOCK	NEW YORK	NY	2122639700	895	99
	f3311f06-425c-4f5b-b544-c5e8e4db36ad	ALFONSO LLOSA GUERRA	NEW YORK	NY	2129391840	895	99
	3882233f-920b-4fd1-b9c8-e66a027e3089	ALICE MIN	NEW YORK	NY	2126598551	895	99
	cb871a66-0414-4008-81f7-39b28a20b7a8	AMY ROSENBERG	NEW YORK	NY	2125236500	895	99
	e427e714-9f1f-4fab-96b4-e7fb7d8046d8	ANDREA HOWARD	NEW YORK	NY	2129391840	895	99
	946c5411-5a3a-4f48-9e70-c828a667048b	ANDY MILLER	NEW YORK	NY	2127747598	895	99
	03ad187b-b228-4db6-b490-596b2434abca	ANGELA GOMEZ-SIMMONDS	NEW YORK	NY	2123059876	895	99
	1ce04851-2a88-4435-b3d7-e023d253d77e	ANNA KALTSAS	NEW YORK	NY	2126392000	895	99
	4f4bad04-f6c5-45d6-8b3e-b942e987f55d	ANNA MARO	NEW YORK	NY	2123265705	895	99
	e22c8e8e-f0ca-4f82-873d-0cb4f03654a9	ANNE-CATRIN UHLEMANN	NEW YORK	NY	2123059876	895	99
	213f03d4-b362-49f4-ac07-4ac24ebb3278	ANTHONY BOWEN	NEW YORK	NY	2123059876	895	99
	e1afa0cd-0d03-480f-98f6-a1d5bc6f9b8a	AYANA MORALES	NEW YORK	NY	2125905152	895	99
	4bf9d55d-af5b-4ae8-9202-c310f6422930	BARRY BRAUSE	NEW YORK	NY	2127747598	895	99
	c847fecc-5292-4143-ab1c-1e887e7e0386	BARRY HARTMAN	NEW YORK	NY	2125905152	895	99

Took about 100 seconds to run (before indexing)

Q3) Coverage gaps: “diseases with no doctors for their specialty in a state”

Use-case: which meds need attention soon because they’re low and scheduled to be taken shortly.

```
WITH state_specs AS (  -- specialties that actually have doctors in
the state
  SELECT DISTINCT d.specialty_id
  FROM doctor d
  WHERE d.state = 'NY'       -- tweak state
),
gap_diseases AS (            -- diseases whose specialty is missing in
the state
  SELECT dis.disease_id, dis.specialty_id
  FROM disease dis
  WHERE NOT EXISTS (
    SELECT 1
    FROM state_specs ss
    WHERE ss.specialty_id = dis.specialty_id
  )
)
SELECT
  gd.disease_id,
  gd.specialty_id,
  COUNT(DISTINCT ds.symptom_id) AS symptom_count  -- “importance”
proxy
FROM gap_diseases gd
JOIN disease_symptom ds ON ds.disease_id = gd.disease_id
GROUP BY gd.disease_id, gd.specialty_id
ORDER BY symptom_count DESC
LIMIT 15;
```

Output:

	disease_id	specialty_id	symptom_count
▶	30fd6064-c5d0-44be-b357-41e9888cf6e6	NULL	9
	2e0033b9-aa90-454e-b5b5-c3a4f1dabb4b	928977a1-89b2-5b65-9ac6-95f29b70316e	6

Took about 110 seconds to run (before indexing)

Q4) Redundancy → “Specialty redundancy score per city”

Use-case: a single endpoint that returns doctors if the user typed a specialty OR described symptoms. You can show the source in results and rank symptom-derived matches higher.

```
WITH city_doctors AS (
  SELECT doctor_id, specialty_id, city, state
  FROM doctor
  WHERE city = 'New York'
),
spec_doctor_counts AS (
  SELECT specialty_id, COUNT(*) AS doctor_count
  FROM city_doctors
  GROUP BY specialty_id
),
spec_disease_counts AS (
  SELECT dis.specialty_id, COUNT(DISTINCT dis.disease_id) AS
disease_count
  FROM disease dis
  GROUP BY dis.specialty_id
)
SELECT
  sdc.specialty_id,
  sdc.doctor_count,
  sdis.disease_count,
  ROUND(sdc.doctor_count / NULLIF(sdis.disease_count,0), 3) AS
doctors_per_disease
FROM spec_doctor_counts sdc
JOIN spec_disease_counts sdis USING (specialty_id)
WHERE sdc.doctor_count >= 1 AND sdis.disease_count >= 1
ORDER BY doctors_per_disease DESC, sdc.doctor_count DESC
LIMIT 15;
```

Output:

specialty_id	doctor_count	disease_count	doctors_per_disease
d33fa24e-00f8-5204-9686-4a724eb31bf6	1490	1	1490.000
89429917-0912-599f-8e29-04af9ab9e46d	207	1	207.000
5812a6ed-a729-5aac-a995-73e0b8f5a550	584	6	97.333
6c7babc9-3ba9-53dc-8ff9-3c3656d1e505	314	5	62.800
5873ac47-a1bf-51c6-8e0f-f35a69dd97c4	104	2	52.000
cb43adca-2acb-5219-ae70-5c0e1bee45b1	326	8	40.750
abc9a45e-b410-5c5b-b867-8b73639bb91e	36	1	36.000
93fb3487-31bf-553b-894c-2fae67e1daf9	107	3	35.667
b620e6b3-bc9e-58a2-a694-5be846d06b35	459	13	35.308
9cf639d8-09ac-5aa7-aa52-3e429987e59a	29	1	29.000
14bd064c-be35-50ba-b3dd-3de84101abd3	215	8	26.875
04c32abc-b8de-5418-8aa8-8f794e60ad18	51	2	25.500
e8582c93-289c-598f-8eb0-3a84fb0b4c0c	49	2	24.500
1d3b58e1-dee9-5a67-80c4-e23720a86366	44	2	22.000
c9ad194f-2ad5-58a7-957a-100a6aee83b9	17	1	17.000

WITH city_doctors AS (SELECT doctor_id, specialty_id, city, state FROM doctor WHERE city = 'New Yo... 15 row(s) returned 103.078 sec / 0.000 sec

Took about 100 seconds to run (before indexing)