

Title:

Influenza (flu) vaccination Data Analysis Project using PostgreSQL.

Requirements:

You will have a few flat files, load them into PostgreSQL and answer the following important healthcare questions.

Problem Statements:

1. Annual Flu Shot Statistics by Demographics:

- How many patients received a flu shot in the year 2022, broken down by age, race, and county?
- Answer query:

Query

Query History

```
1 SELECT
2     p.race,
3     p.county,
4     CASE
5         WHEN EXTRACT(YEAR FROM AGE('2024-01-01', p.BIRTHDATE)) < 18 THEN '0-17'
6         WHEN EXTRACT(YEAR FROM AGE('2024-01-01', p.BIRTHDATE)) BETWEEN 18 AND 35 THEN '18-35'
7         WHEN EXTRACT(YEAR FROM AGE('2024-01-01', p.BIRTHDATE)) BETWEEN 36 AND 59 THEN '36-59'
8         ELSE '60+'
9     END AS AgeGroup,
10    COUNT(*) AS GroupCount
11 FROM patients as p
12 JOIN immunizations as i ON p.id = i.patient
13 WHERE EXTRACT(YEAR FROM i.date) = 2022
14        AND i.description LIKE '%Flu%'
15 GROUP BY p.race, p.county, AgeGroup, i.description
16
```

Data Output

Messages

Notifications

	race character varying (100)	county character varying (100)	agegroup text	groupcount bigint
1	asian	Barnstable County	0-17	1
2	asian	Barnstable County	60+	1
3	asian	Berkshire County	18-35	1
4	asian	Berkshire County	36-59	4
5	asian	Berkshire County	60+	1
6	asian	Bristol County	0-17	5
7	asian	Bristol County	18-35	3
8	asian	Bristol County	36-59	7
9	asian	Bristol County	60+	4
10	asian	Dukes County	36-59	2
11	asian	Essex County	0-17	10
12	asian	Essex County	18-35	12
13	asian	Essex County	36-59	11
14	asian	Essex County	60+	11
15	asian	Hampden County	18-35	2
16	asian	Hampden County	36-59	5

Total rows: 233 of 233

Query complete 00:00:00.132

2. Percentage Analysis of Flu Shots:

- What is the percentage of patients who received a flu shot in 2022 compared to the total number of patients, stratified by age, race, and county?
- Answer query:

```

8 ----- Below is the answer of question 2 (a)
9 SELECT
10     p.race,
11     p.county,
12     EXTRACT(YEAR FROM AGE('2024-01-01', p.birthdate)) AS Age
13 FROM
14     patients as p
15 JOIN
16     immunizations as i
17 ON
18     p.Id = i.patient
19 WHERE
20     i.description LIKE '%Flu%'
21     AND EXTRACT(YEAR FROM i.date) = 2022
22 GROUP BY
23     p.race,
24     p.county,
25     Age
26

```

Data Output Messages Notifications

	race character varying (100)	county character varying (100)	age numeric
1	asian	Barnstable County	4
2	asian	Barnstable County	74
3	asian	Berkshire County	20
4	asian	Berkshire County	37
5	asian	Berkshire County	39
6	asian	Berkshire County	59
7	asian	Berkshire County	62
8	asian	Bristol County	3
9	asian	Bristol County	9
10	asian	Bristol County	13
11	asian	Bristol County	14
12	asian	Bristol County	17
13	asian	Bristol County	19
14	asian	Bristol County	23
Total rows: 1000 of 1983			Query complete 00:00:00.134

- What is the overall percentage of patients who received a flu shot in 2022 across the entire hospital or clinic?
- Answer query:

```

36 ----- Below is the answer of question 2 (b)
37 SELECT
38     (SELECT COUNT(i.patient) --specific
39     FROM immunizations as i
40     WHERE i.description LIKE '%Flu%'
41     AND EXTRACT(YEAR FROM i.date) = 2022) * 100 /
42     (SELECT COUNT(i.patient) --total
43     FROM immunizations as i) AS percentage_of_vaccinated
44
45

```

Data Output Messages Notifications

	percentage_of_vaccinated bigint
1	5

3. Monthly Cumulative Flu Shot Data:

- How many flu shots were administered each month during the year 2022?
- Answer query:

Query

Query History

```
1 SELECT COUNT(patient)
2 FROM immunizations as i
3 WHERE i.description LIKE '%Flu%'
4       AND EXTRACT(YEAR FROM i.date) = 2022
5 -- 8437 flu shots were administered in the year 2020.
6
7 -- Now lets calculate flu shots per month
8 SELECT
9     SUM(CASE WHEN EXTRACT(MONTH FROM i.date) = 1 THEN 1 ELSE 0 END) AS Jan,
10    SUM(CASE WHEN EXTRACT(MONTH FROM i.date) = 2 THEN 1 ELSE 0 END) AS Feb,
11    SUM(CASE WHEN EXTRACT(MONTH FROM i.date) = 3 THEN 1 ELSE 0 END) AS Mar,
12    SUM(CASE WHEN EXTRACT(MONTH FROM i.date) = 4 THEN 1 ELSE 0 END) AS Apr,
13    SUM(CASE WHEN EXTRACT(MONTH FROM i.date) = 5 THEN 1 ELSE 0 END) AS May,
14    SUM(CASE WHEN EXTRACT(MONTH FROM i.date) = 6 THEN 1 ELSE 0 END) AS Jun,
15    SUM(CASE WHEN EXTRACT(MONTH FROM i.date) = 7 THEN 1 ELSE 0 END) AS Jul,
16    SUM(CASE WHEN EXTRACT(MONTH FROM i.date) = 8 THEN 1 ELSE 0 END) AS Aug,
17    SUM(CASE WHEN EXTRACT(MONTH FROM i.date) = 9 THEN 1 ELSE 0 END) AS Sep,
18    SUM(CASE WHEN EXTRACT(MONTH FROM i.date) = 10 THEN 1 ELSE 0 END) AS Oct,
19    SUM(CASE WHEN EXTRACT(MONTH FROM i.date) = 11 THEN 1 ELSE 0 END) AS Nov,
20    SUM(CASE WHEN EXTRACT(MONTH FROM i.date) = 12 THEN 1 ELSE 0 END) AS Dec
21 FROM
22     immunizations as i
23 WHERE
24     i.description LIKE '%Flu%'
25     AND EXTRACT(YEAR FROM i.date) = 2022
```

Data Output

Messages

Graph Visualiser

×

Notifications

jan

feb

mar

apr

may

jun

jul

aug

sep

oct

nov

dec

bigint

bigint

bigint

bigint

bigint

bigint

bigint

bigint

bigint

bigint

bigint

bigint

1

656

616

757

892

723

661

712

711

669

704

667

669

4. Total Annual Flu Shots:

- What is the total number of flu shots given in the year 2022?
- Answer query:

Query

Query History

```
1 SELECT COUNT(patient)
2 FROM immunizations as i
3 WHERE i.description LIKE '%Flu%'
4       AND EXTRACT(YEAR FROM i.date) = 2022
5 -- 8437 flu shots were administered in the year 2020.
6
```

Data Output

Messages

Graph Visualiser

×

Notifications

≡

+

▼

▼

▼

▼

▼

▼

count

bigint

1

8437

5. Patient Lists for Flu Shot Analysis:

- Can you provide a list of patients who received a flu shot in 2022?
- Answer query:

```

1 SELECT i.patient, p.first, p.last, p.gender, p.county, i.description
2 FROM immunizations as i
3 JOIN patients as p
4 ON i.patient = p.Id
5 WHERE i.description LIKE '%Flu%'
6     AND EXTRACT(YEAR FROM i.date) = 2022
7 -- list of patients who received a flu shot in 2022

```

Data Output

Messages

Notifications

patient character varying (100)	first character varying (100)	last character varying (100)	gender character varying (100)	county character varying (100)	description character varying (500)
1 73d3e63-e656-b9de-fd61-88d370a86d...	Hedy	Von	F	Middlesex County	Seasonal Flu Vaccine
2 32a2188a-132e-4fd4-1a0e-3d532f4c4ea8	Laurn	Wisozk	F	Suffolk County	Seasonal Flu Vaccine
3 4adcad4e-1aa5-5601-4107-5595314847...	Dorthea	Reichel	F	Norfolk County	Seasonal Flu Vaccine
4 52de8610-de40-203d-4bd6-cf06141fa864	Sherita	Orm	F	Hampden County	Seasonal Flu Vaccine
5 ac63da4a-893a-1d3c-aa93-fb78c8da3d95	Casie	Hilpert	F	Norfolk County	Seasonal Flu Vaccine
6 10d940ae-7114-8bcf-50f9-2bfd5354ff41	Rebeca	Mayer	F	Essex County	Seasonal Flu Vaccine
7 154290c8-6729-fe33-d3b6-a66f04bb939e	Elenora	Raynor	F	Barnstable County	Seasonal Flu Vaccine
8 91f57fe7-7176-6f23-1c4b-eb973e556ce3	Rosario	Ruecker	F	Norfolk County	Seasonal Flu Vaccine

- Can you also provide a list of patients who did not receive a flu shot in 2022?
- Answer query:

Query

Query History

```
9 SELECT i.patient, p.first, p.last, p.gender, p.county, i.description
10 FROM immunizations as i
11 JOIN patients as p
12 ON i.patient = p.id
13 WHERE i.description NOT LIKE '%Flu%'
14 AND EXTRACT(YEAR FROM i.date) = 2022
15 -- list of patients who did not receive a flu shot in 2022
```

Data Output

Messages

Notifications

	patient character varying (100)	first character varying (100)	last character varying (100)	gender character varying (100)	county character varying (100)	description character varying (500)
1	ac53da4a-893a-1d3c-aa93-fb78c8da3d95	Casie	Hilpert	F	Norfolk County	Herpes Zoster Vaccine (Live)
2	ac53da4a-893a-1d3c-aa93-fb78c8da3d95	Casie	Hilpert	F	Norfolk County	Five doses of tetanus toxoid, preservative-free and adsorbed, for adul...
3	6cc89ec7-8e4a-536a-556a-05689f3fac7a	Dakota	Kihn	F	Worcester County	Novel Coronavirus (COVID-19) mRNA Vaccine 30 mcg/0.3mL Dose
4	6cc89ec7-8e4a-536a-556a-05689f3fac7a	Dakota	Kihn	F	Worcester County	Human Papillomavirus (HPV) Four-strain Vaccine
5	6cc89ec7-8e4a-536a-556a-05689f3fac7a	Dakota	Kihn	F	Worcester County	Novel Coronavirus (COVID-19) mRNA Vaccine 30 mcg/0.3mL Dose
6	3a6e47b6-2002-21ee-a8d1-4fa836150a1	Lissa	Luetlgen	F	Plymouth County	Novel Coronavirus (COVID-19) mRNA Vaccine 30 mcg/0.3mL Dose
7	3124c0fc-fa3a-480e-bd23-d7c89c099e51	Lorlee	Kemmer	F	Barnstable County	Five doses of Tetanus toxoid, preservative-free and adsorbed, for adul...
8	c216356a-fe02-9b37-2a40-f130541e8636	Colene	Bailey	F	Essex County	Five doses of Tetanus toxoid, preservative-free and adsorbed, for adul...

More Questions that should be answered:

- Patient Demographics Analysis: How many patients are in each age group (e.g., 0-18, 19-35, 36-60, 60+)?

Query

Query History

```
1 -- 1. Patient Demographics Analysis:
2 -- How many patients are in each age group (e.g., 0-18, 19-35, 36-60, 60+)?
3 SELECT
4 CASE
5     WHEN EXTRACT(YEAR FROM AGE('2024-01-01', p.birthdate)) <= 18 THEN '0-18'
6     WHEN EXTRACT(YEAR FROM AGE('2024-01-01', p.birthdate)) BETWEEN 19 and 35 THEN '19-35'
7     WHEN EXTRACT(YEAR FROM AGE('2024-01-01', p.birthdate)) BETWEEN 36 and 60 THEN '36-60'
8     ELSE '60+'
9 END AS AgeGroup,
10 COUNT(*) AS GroupCount
11 FROM patients as p
12 GROUP BY AgeGroup
```

Data Output

Messages

Notifications

	agegroup text	groupcount bigint
1	19-35	2413
2	60+	3183
3	36-60	3661
4	0-18	2106

- Geographical Distribution: What are the top 5 cities with the highest number of patients?

```

14 -- 2. Geographical Distribution:
15 -- What are the top 5 cities with the highest number of patients?
16 SELECT p.city, COUNT(p.id)
17 FROM patients as p
18 GROUP BY p.city
19 ORDER BY COUNT(p.id) DESC
20 LIMIT 5

```

Data Output Messages Notifications		
	city character varying (100)	count bigint
1	Boston	1016
2	Worcester	295
3	Springfield	244
4	Lowell	193
5	Cambridge	166

3. Healthcare Utilization Patterns: How many encounters does each patient have on average?

1. There are 2 meanings of the question: I've solved both.

1. This means that how many encounters does each patient have:

Query Query History		
<pre> 22 -- 3. Healthcare Utilization Patterns: 23 -- How many encounters does each patient have on average? 24 SELECT patient, COUNT(*) AS encounter_count 25 FROM conditions 26 GROUP BY patient; </pre>		
Data Output Messages Notifications		
	patient character varying (1000)	encounter_count bigint
1	519816e7-29d9-dc7e-00ba-cf949569cb...	5
2	f118199c-a128-e36a-dc75-a5699ccf94f7	14
3	1b5d0889-ce72-c592-147f-fe2596d693...	18
4	009bdb22-c99b-2a3d-a907-3d30d5a7a...	1
5	3e123fca-8168-4a02-b309-e3cf8de0b6...	18

2. This shows how many encounters does each patient have on average, so the answer will be just one for all patients and not different for each patient:

Query Query History		
<pre> 29 SELECT 30 AVG(encounter_count) AS AverageEncounters 31 FROM (32 SELECT 33 patient, 34 COUNT(*) AS encounter_count 35 FROM 36 encounters 37 GROUP BY 38 patient 39); </pre>		
Data Output Messages Notifications		
	averageencounters numeric	
1	43.5967680244788679	

4. Condition Prevalence: What are the top 10 most common conditions among patients?

```

43 -- 4. Condition Prevalence:
44 -- What are the top 10 most common conditions among patients?
45 SELECT description, COUNT(*) AS MostCommon
46 FROM conditions
47 GROUP BY description
48 ORDER BY MostCommon desc
49 LIMIT 10

```

Data Output Messages Notifications		
	description character varying (200)	mostcommon bigint
1	Other psychological or physical stress, not elsewhere classifi...	58962
2	Pregnant state, incidental	9872
3	Acute bronchitis	5684
4	Body Mass Index 30.0-30.9, adult	5461
5	Unemployment	4717
6	Anemia, unspecified	3704
7	Other specified anemias	3704
8	Unspecified essential hypertension	3682
9	Other chronic pain	2727
10	Tubal ligation status	2537

5. Healthcare Costs Analysis: What is the average healthcare expense per patient?

```

51 -- 5. Healthcare Costs Analysis:
52 -- What is the average healthcare expense per patient?
53 SELECT avg(healthcare_expenses)
54 FROM patients

```

Data Output Messages Graph Visualiser Notifications		
	avg double precision	
1	272690.2093751649	

6. Encounter Analysis: Which type of encounters (as per ENCOUNTERCLASS) is most common?

```

56 -- 6. Encounter Analysis:
57 -- Which type of encounters (as per ENCOUNTERCLASS) is most common?
58 SELECT encounterclass, COUNT(encounterclass) AS Count_of_each
59 FROM encounters
60 GROUP BY encounterclass
61 ORDER BY Count_of_each desc

```

Data Output Messages Notifications		
	encounterclass character varying (100)	count_of_each bigint
1	ambulatory	244148
2	outpatient	85849
3	wellness	77849
4	urgentcare	21479
5	emergency	17903
6	inpatient	3709
7	home	2318
8	snf	1221
9	hospice	750
10	virtual	709

7. Income vs. Healthcare Coverage: Is there a correlation between patients' income levels and their healthcare coverage amounts?

```

63 -- 7. Income vs. Healthcare Coverage:
64 -- Is there a correlation between patients' income levels and their healthcare coverage amounts?
65 SELECT
66     IncomeBracket,
67     AVG(healthcare_coverage) AS AverageCoverage
68 FROM
69     (SELECT
70         healthcare_coverage,
71         CASE
72             WHEN income <= 1000 THEN '0-1K'
73             WHEN income > 1000 AND income <= 10000 THEN '1K-10K'
74             WHEN income > 10000 AND income <= 50000 THEN '10K-50K'
75             WHEN income > 50000 AND income <= 100000 THEN '50K-100K'
76             WHEN income > 100000 AND income <= 500000 THEN '100K-500K'
77             WHEN income > 500000 THEN '500K+'
78             ELSE 'Unknown'
79         END AS IncomeBracket
80     FROM
81         patients) AS IncomeBracketed
82 GROUP BY
83     IncomeBracket
84 ORDER BY
85     CASE IncomeBracket
86         WHEN '0-1K' THEN 1
87         WHEN '1K-10K' THEN 2
88         WHEN '10K-50K' THEN 3
89         WHEN '50K-100K' THEN 4
90         WHEN '100K-500K' THEN 5
91         WHEN '500K+' THEN 6
92         ELSE 7
93     END

```

incomebracket	averagecoverage
0-1K	490905.818363644
1K-10K	824126.4631939803
10K-50K	627516.2361551143
50K-100K	383668.8178207116
100K-500K	573441.8988333847
500K+	622949.2446915581

8. Immunization Records: How many patients have received each type of immunization?

```

95 -- 8. Immunization Records:
96 -- How many patients have received each type of immunization?
97 SELECT description, COUNT(patient) as Number_of_Patients
98 FROM immunizations
99 GROUP BY description
100 ORDER BY COUNT(patient) desc

```

description	number_of_patients
Seasonal Flu Vaccine	93219
Five doses of tetanus toxoid, preservative-free and adsorbed, for adults.	8434
Novel Coronavirus (COVID-19) mRNA Vaccine 30 mcg/0.3mL Dose	7563
Diphtheria, Tetanus, and Pertussis Vaccine	6693
Novel Coronavirus (COVID-19) mRNA Vaccine 100 mcg/0.5mL Dose	5993

9. Patient Mortality: What is the average age at death for patients, and how does it vary by gender and race?

```

102 -- 9. Patient Mortality:
103 -- What is the average age at death for patients, and how does it vary by gender and race?
104 SELECT gender, race,
105     AVG(EXTRACT(YEAR FROM AGE(patients.deathdate, patients.birthdate))) AS Age_of_patient
106 FROM patients
107 WHERE deathdate IS NOT NULL
108     AND birthdate IS NOT NULL
109 GROUP BY race, gender

```

gender	race	age_of_patient
F	hawaiian	61.6896551724137931
F	asian	55.6172839506172840
F	native	40.1250000000000000
F	black	56.7402597402597403
F	white	53.4388224471021159
F	other	43.6000000000000000

10. Provider Workload: Which healthcare provider has attended the most encounters?

```

111 -- 10. Provider Workload:
112 -- Which healthcare provider has attended the most encounters?
113 SELECT provider, COUNT(provider) AS count_encounter
114 FROM encounters
115 GROUP BY provider
116 ORDER BY count_encounter desc
117 LIMIT 1

```

Data Output Messages Notifications		
	provider character varying (100)	count_encounter bigint
1	4e98e792-2919-3258-9159-025edd33f95c	13470

11. Encounter Costs Analysis: What are the average total claim costs and payer coverage for each encounter class?

```

119 -- 11. Encounter Costs Analysis:
120 -- What are the average total claim costs and payer coverage for each encounter class?
121 SELECT encounterclass, AVG(total_claim_cost) as average_claim, AVG(payer_coverage) as average_coverage
122 FROM encounters
123 GROUP BY encounterclass

```

Data Output Messages Graph Visualiser X Notifications			
	encounterclass character varying (100)	average_claim double precision	average_coverage double precision
1	ambulatory	4910.160586775049	4060.540984484715
2	emergency	3876.8564486398027	3052.0752583365756
3	home	692.5313028472748	566.3527610008498
4	hospice	11124.481733333325	9135.34462666667
5	inpatient	17915.354041520615	14461.583362092226
6	outpatient	1986.2282548432086	1188.7972674113473
7	snf	15394.501990171957	11386.712850122847
8	urgentcare	1034.4758773686017	787.8173932678461
9	virtual	981.3493511988781	868.4101833568405
10	wellness	1124.008335880981	534.0480825700955

12. Marital Status and Health: Is there a notable difference in the number of healthcare encounters between married and single patients?

```

125 -- 12. Marital Status and Health:
126 -- Is there a notable difference in the number of healthcare encounters between married and single patients?
127 SELECT p.marital, COUNT(e.id) AS NumberOfEncounters
128 FROM patients AS p
129 JOIN encounters AS e
130 ON p.id = e.patient
131 WHERE p.marital IS NOT NULL
132 AND (p.marital = 'S' OR p.marital = 'M')
133 GROUP BY p.marital

```

Data Output Messages Graph Visualiser X Notifications		
	marital character varying (100)	numberofencounters bigint
1	M	218310
2	S	75780

13. Race and Ethnicity in Healthcare: How do healthcare expenses vary across different races and ethnicities?


```

135 -- 13. Race and Ethnicity in Healthcare:
136 -- How do healthcare expenses vary across different races and ethnicities?
137 SELECT p.race, AVG(p.healthcare_expenses) AS avg_healthcare_exp, p.ethnicity
138 FROM patients AS p
139 WHERE p.race IS NOT NULL
140       AND p.healthcare_expenses IS NOT NULL
141       AND p.ethnicity IS NOT NULL
142 GROUP BY p.race, p.ethnicity
143 ORDER BY avg_healthcare_exp desc

```

	race character varying (100)	avg_healthcare_exp double precision	ethnicity character varying (100)
1	black	305348.6220138888	nonhispanic
2	hawaiian	292749.60713178304	nonhispanic
3	other	288273.9413402062	nonhispanic
4	black	282769.26225806447	hispanic
5	white	274622.1006352431	nonhispanic
6	hawaiian	256504.12083333332	hispanic
7	asian	255636.98189189183	hispanic
8	asian	247047.32128637074	nonhispanic
9	white	246105.73668999188	hispanic
10	native	212815.9646808511	nonhispanic
11	native	164523.97500000003	hispanic
12	other	126811.84500000002	hispanic

14. Chronic Condition Management: Which patients with chronic conditions (conditions lasting more than 1 year) have the highest number of encounters?

```

145 -- 14. Chronic Condition Management:
146 -- Which patients with chronic conditions (conditions lasting more than 1 year) have the highest number of encounters?
147 SELECT
148     EXTRACT(YEAR FROM AGE(c.stop, c.start)) AS Chronic_years,
149     c.patient,
150     c.encounter
151 FROM conditions AS c
152 WHERE c.start IS NOT NULL
153       AND c.stop IS NOT NULL
154       AND c.patient IS NOT NULL
155       AND c.encounter IS NOT NULL
156       AND EXTRACT(YEAR FROM AGE(c.stop, c.start)) >= 1
157 ORDER BY EXTRACT(YEAR FROM AGE(c.stop, c.start)) DESC;

```

	chronic_years numeric	patient character varying (1000)	encounter character varying (1000)
1	52	bad6ff27-8229-ca89-fd82-10b8c93cc9bc	6bf53a35-752a-d260-1a6e-0dcfd1b01cb
2	51	5e0c39ce-38f7-d66b-aaac-0c61edfe8859	352484c5-1dab-49ab-dbec-5356d45b07...
3	48	13fa35d6-3006-417d-e598-3c03da223c73	48f8bbfc-f52c-6798-8237-2475193682e9

15. Patient Coverage Gaps: Identify patients with high healthcare expenses but low healthcare coverage.

```

159 -- 15. Patient Coverage Gaps:
160 -- Identify patients with high healthcare expenses but low healthcare coverage.
161 SELECT id,
162        healthcare_expenses,
163        healthcare_coverage
164 FROM patients
165 WHERE
166     healthcare_expenses > (SELECT AVG(healthcare_expenses) FROM patients) -- Above average expenses
167     AND healthcare_coverage < (SELECT AVG(healthcare_coverage) FROM patients) -- Below average coverage
168

```

	id character varying (100)	healthcare_expenses double precision	healthcare_coverage double precision
1	cc53e99a-6715-603f-b074-eea93fe11e20	824192.36	31979.82
2	52de8610-de40-203d-4bd6-cf06141fa864	556114.94	11443.34
3	10d940ae-7114-8bcf-50f9-2bfd5354ff41	654622.19	270730.9