**Project 1:** Synthetic Clinical Patient Database

Team G66: Allie Schneider, Francis Villamater, Andrew Chen

**Overall instructions**

In this milestone, you will review the data design regarding birthplace and living place:

* The new data model uses Z\_\* tables instead of N\_\* tables
* You must reproduce the task required in Milestone 3 (ingest data from T\_ tables into Z\_ tables), but now using the new proposed data model (differences in the way city/county/state/country are designed).
  + **DO NOT use/reuse any of the N\_ tables in milestone #4**
* Then work in the additional tables (allergy)
* Your submission MUST include (Word file):
* This document only, including DML statements (NOT as screenshot), and screenshots in each question to show the result of each one

**1. INGEST DATA**

You must present all the required DML statements (INSERT, UPDATE, DELETE) used to populate all Z\_ tables (show them in the correct order). Screenshots with DML statements are NOT allowed.

You DON’T need to provide the DDL used to create tables.

**\*\*\* DO NOT CHANGE data in T\_PATIENT \*\*\***

**2. QUERIES: answer the questions below using T\_ and Z\_ tables only, but separately (not N\_ tables)**

|  |  |
| --- | --- |
| Question 1: | **What is the distribution of patients regarding birthplace (country/state/city)?** |
| T Query (SQL): | SELECT BIRTH\_PLACE, COUNT(\*)  FROM T\_PATIENT  GROUP BY BIRTH\_PLACE  ORDER BY COUNT(\*) DESC; |
| T Result: |  |
| Z Query (SQL): | SELECT BIRTH\_PLACE\_COUNTRY\_ID, BIRTH\_PLACE\_STATE\_ID, BIRTH\_PLACE\_CITY\_ID, COUNT(\*)  FROM Z\_PATIENT  GROUP BY BIRTH\_PLACE\_COUNTRY\_ID, BIRTH\_PLACE\_STATE\_ID, BIRTH\_PLACE\_CITY\_ID  ORDER BY COUNT(\*) DESC; |
| Z Result: |  |
| Comments | Any differences on the two results?  There distributions across both queries are the same. |

|  |  |
| --- | --- |
| Question 2: | **How many patients are missing information about living county?** |
| T Query (SQL): | SELECT COUNT(\*)  FROM T\_PATIENT  WHERE COUNTY IS NULL; |
| T Result: |  |
| Z Query (SQL): | SELECT COUNT(\*)  FROM Z\_PATIENT a  JOIN Z\_CITY b ON a.LIVING\_PLACE\_CITY\_ID = b.CITY\_ID  WHERE b.COUNTY\_ID IS NULL; |
| Z Result: |  |
| Comments | Any differences on the two results?  There is no difference in the two results. |

|  |  |
| --- | --- |
| Question 3: | **What is the allergy affecting the greatest number of patients?** |
| T Query (SQL): | SELECT ALLERGY\_DESCRIPTION, COUNT(\*)  FROM T\_PATIENT\_ALLERGY  GROUP BY ALLERGY\_DESCRIPTION  ORDER BY COUNT(\*) DESC; |
| T Result: |  |
| Z Query (SQL): | SELECT b.ALLERGY\_DESCRIPTION, COUNT(\*)  FROM Z\_PATIENT\_ALLERGY a  JOIN Z\_ALLERGY b ON a.ALLERGY\_CODE = b.ALLERGY\_CODE  GROUP BY b.ALLERGY\_DESCRIPTION  ORDER BY COUNT(\*) DESC; |
| Z Result: |  |
| Comments | Any differences on the two results?  There are no differences in the two results. |

|  |  |
| --- | --- |
| Question 4: | **How is that same allergy (from the previous questions) affecting distinct counties regarding number of patients affected?** |
| T Query (SQL): | SELECT a.COUNTY, COUNT(\*)  FROM T\_PATIENT a  JOIN T\_PATIENT\_ALLERGY b ON a.PATIENT\_ID = b.PATIENT\_ID  WHERE b.ALLERGY\_DESCRIPTION = 'Allergy to mould'  GROUP BY a.COUNTY  ORDER BY COUNT(\*) DESC; |
| T Result: |  |
| Z Query (SQL): | SELECT \*  FROM Z\_ALLERGY  WHERE ALLERGY\_DESCRIPTION = 'Allergy to mould';  SELECT b.COUNTY\_ID, COUNT(\*)  FROM Z\_PATIENT a  JOIN Z\_CITY b ON a.LIVING\_PLACE\_CITY\_ID = b.CITY\_ID  JOIN Z\_PATIENT\_ALLERGY c ON a.PATIENT\_ID = c.PATIENT\_ID  WHERE c.ALLERGY\_CODE = '419474003'  GROUP BY b.COUNTY\_ID  ORDER BY COUNT(\*) DESC; |
| Z Result: |  |
| Comments | Any differences on the two results?  There are no differences in the two results. |

|  |  |
| --- | --- |
| Question 5: | **Which of the top-5 allergies affecting number of male patients are also in the list of top-5 allergies affecting number of female patients (top-5 overlapping)?** |
| T Query (SQL): | SELECT a.ALLERGY\_CODE  FROM (SELECT a.ALLERGY\_CODE, COUNT(\*)  FROM T\_PATIENT\_ALLERGY a  JOIN T\_PATIENT b ON a.PATIENT\_ID = b.PATIENT\_ID  WHERE b.GENDER = 'M'  GROUP BY a.ALLERGY\_CODE  ORDER BY COUNT(\*) DESC  FETCH FIRST 5 ROWS ONLY) a  INNER JOIN (SELECT a.ALLERGY\_CODE, COUNT(\*)  FROM T\_PATIENT\_ALLERGY a  JOIN T\_PATIENT b ON a.PATIENT\_ID = b.PATIENT\_ID  WHERE b.GENDER = 'F'  GROUP BY a.ALLERGY\_CODE  ORDER BY COUNT(\*) DESC  FETCH FIRST 5 ROWS ONLY) b ON a.ALLERGY\_CODE = b.ALLERGY\_CODE; |
| T Result: |  |
| Z Query (SQL): | SELECT a.ALLERGY\_CODE  FROM (SELECT a.ALLERGY\_CODE, COUNT(\*)  FROM Z\_PATIENT\_ALLERGY a  JOIN Z\_PATIENT b ON a.PATIENT\_ID = b.PATIENT\_ID  WHERE b.GENDER\_ID = 'M'  GROUP BY a.ALLERGY\_CODE  ORDER BY COUNT(\*) DESC  FETCH FIRST 5 ROWS ONLY) a  INNER JOIN (SELECT a.ALLERGY\_CODE, COUNT(\*)  FROM Z\_PATIENT\_ALLERGY a  JOIN Z\_PATIENT b ON a.PATIENT\_ID = b.PATIENT\_ID  WHERE b.GENDER\_ID = 'F'  GROUP BY a.ALLERGY\_CODE  ORDER BY COUNT(\*) DESC  FETCH FIRST 5 ROWS ONLY) b ON a.ALLERGY\_CODE = b.ALLERGY\_CODE; |
| Z Result: |  |
| Comments | Any differences on the two results?  There is no difference in the two results. |

**Challenging questions**:

|  |  |
| --- | --- |
| Challenge 1: | **Which living cities have the biggest difference regarding how allergies (in general) proportionally affect each gender?** |
| T Query (SQL): | SELECT a.CITY, ABS(a.MALE\_ALLERGY\_COUNT - b.FEMALE\_ALLERGY\_COUNT) AS ALLERGY\_DIFFERENCE  FROM (SELECT a.CITY, COUNT(b.ALLERGY\_CODE) AS MALE\_ALLERGY\_COUNT  FROM T\_PATIENT a  JOIN T\_PATIENT\_ALLERGY b ON a.PATIENT\_ID = b.PATIENT\_ID  WHERE a.GENDER = 'M'  GROUP BY a.CITY  ORDER BY COUNT(b.ALLERGY\_CODE) DESC) a  JOIN (SELECT a.CITY, COUNT(b.ALLERGY\_CODE) AS FEMALE\_ALLERGY\_COUNT  FROM T\_PATIENT a  JOIN T\_PATIENT\_ALLERGY b ON a.PATIENT\_ID = b.PATIENT\_ID  WHERE a.GENDER = 'F'  GROUP BY a.CITY  ORDER BY COUNT(b.ALLERGY\_CODE) DESC) b ON a.CITY = b.CITY  ORDER BY ALLERGY\_DIFFERENCE DESC; |
| Result: |  |
| Z Query (SQL): | SELECT a.LIVING\_PLACE\_CITY\_ID, ABS(a.MALE\_ALLERGY\_COUNT - b.FEMALE\_ALLERGY\_COUNT) AS ALLERGY\_DIFFERENCE  FROM (SELECT a.LIVING\_PLACE\_CITY\_ID, COUNT(b.ALLERGY\_CODE) AS MALE\_ALLERGY\_COUNT  FROM Z\_PATIENT a  JOIN Z\_PATIENT\_ALLERGY b ON a.PATIENT\_ID = b.PATIENT\_ID  WHERE a.GENDER\_ID = 'M'  GROUP BY a.LIVING\_PLACE\_CITY\_ID  ORDER BY COUNT(b.ALLERGY\_CODE) DESC) a  JOIN (SELECT a.LIVING\_PLACE\_CITY\_ID, COUNT(b.ALLERGY\_CODE) AS FEMALE\_ALLERGY\_COUNT  FROM Z\_PATIENT a  JOIN Z\_PATIENT\_ALLERGY b ON a.PATIENT\_ID = b.PATIENT\_ID  WHERE a.GENDER\_ID = 'F'  GROUP BY a.LIVING\_PLACE\_CITY\_ID  ORDER BY COUNT(b.ALLERGY\_CODE) DESC) b ON a.LIVING\_PLACE\_CITY\_ID = b.LIVING\_PLACE\_CITY\_ID  ORDER BY ALLERGY\_DIFFERENCE DESC; |
| Z Result: |  |
| Comments | Any differences on the two results?  There are no differences on the two results. |

|  |  |
| --- | --- |
| Challenge 2: | **What is the number of patients affected by each allergy, showing in each column (expected 12 columns in this result):**   1. **Allergy Name** 2. **Total number of patients** 3. **Male (number of)** 4. **Female (number of)** 5. **Single (number of)** 6. **Married (number of)** 7. **White (number of)** 8. **Black (number of)** 9. **Asian (number of)** 10. **Native (number of)** 11. **Hispanic (number of)** 12. **Non-Hispanic (number of)** |
| T Query (SQL): | SELECT a.ALLERGY\_DESCRIPTION, a.ALL\_COUNT, b.MALE\_COUNT, c.FEMALE\_COUNT, d.SINGLE\_COUNT,  e.MARRIED\_COUNT, f.WHITE\_COUNT, g.BLACK\_COUNT, h.ASIAN\_COUNT, i.NATIVE\_COUNT,  j.HISPANIC\_COUNT, k.NONHISPANIC\_COUNT  FROM (SELECT a.ALLERGY\_DESCRIPTION, COUNT(\*) AS ALL\_COUNT  FROM T\_PATIENT\_ALLERGY a  JOIN T\_PATIENT b ON a.PATIENT\_ID = b.PATIENT\_ID  GROUP BY a.ALLERGY\_DESCRIPTION  ORDER BY a.ALLERGY\_DESCRIPTION) a  JOIN (SELECT a.ALLERGY\_DESCRIPTION, COUNT(\*) AS MALE\_COUNT  FROM T\_PATIENT\_ALLERGY a  JOIN T\_PATIENT b ON a.PATIENT\_ID = b.PATIENT\_ID  WHERE b.GENDER = 'M'  GROUP BY a.ALLERGY\_DESCRIPTION  ORDER BY a.ALLERGY\_DESCRIPTION) b ON a.ALLERGY\_DESCRIPTION = b.ALLERGY\_DESCRIPTION  JOIN (SELECT a.ALLERGY\_DESCRIPTION, COUNT(\*) AS FEMALE\_COUNT  FROM T\_PATIENT\_ALLERGY a  JOIN T\_PATIENT b ON a.PATIENT\_ID = b.PATIENT\_ID  WHERE b.GENDER = 'F'  GROUP BY a.ALLERGY\_DESCRIPTION  ORDER BY a.ALLERGY\_DESCRIPTION) c ON a.ALLERGY\_DESCRIPTION = c.ALLERGY\_DESCRIPTION  JOIN (SELECT a.ALLERGY\_DESCRIPTION, COUNT(\*) AS SINGLE\_COUNT  FROM T\_PATIENT\_ALLERGY a  JOIN T\_PATIENT b ON a.PATIENT\_ID = b.PATIENT\_ID  WHERE b.MARITAL\_STATUS = 'S'  GROUP BY a.ALLERGY\_DESCRIPTION  ORDER BY a.ALLERGY\_DESCRIPTION) d ON a.ALLERGY\_DESCRIPTION = d.ALLERGY\_DESCRIPTION  JOIN (SELECT a.ALLERGY\_DESCRIPTION, COUNT(\*) AS MARRIED\_COUNT  FROM T\_PATIENT\_ALLERGY a  JOIN T\_PATIENT b ON a.PATIENT\_ID = b.PATIENT\_ID  WHERE b.MARITAL\_STATUS = 'M'  GROUP BY a.ALLERGY\_DESCRIPTION  ORDER BY a.ALLERGY\_DESCRIPTION) e ON a.ALLERGY\_DESCRIPTION = e.ALLERGY\_DESCRIPTION  JOIN (SELECT a.ALLERGY\_DESCRIPTION, COUNT(\*) AS WHITE\_COUNT  FROM T\_PATIENT\_ALLERGY a  JOIN T\_PATIENT b ON a.PATIENT\_ID = b.PATIENT\_ID  WHERE b.RACE = 'white'  GROUP BY a.ALLERGY\_DESCRIPTION  ORDER BY a.ALLERGY\_DESCRIPTION) f ON a.ALLERGY\_DESCRIPTION = f.ALLERGY\_DESCRIPTION  JOIN (SELECT a.ALLERGY\_DESCRIPTION, COUNT(\*) AS BLACK\_COUNT  FROM T\_PATIENT\_ALLERGY a  JOIN T\_PATIENT b ON a.PATIENT\_ID = b.PATIENT\_ID  WHERE b.RACE = 'black'  GROUP BY a.ALLERGY\_DESCRIPTION  ORDER BY a.ALLERGY\_DESCRIPTION) g ON a.ALLERGY\_DESCRIPTION = g.ALLERGY\_DESCRIPTION  JOIN (SELECT a.ALLERGY\_DESCRIPTION, COUNT(\*) AS ASIAN\_COUNT  FROM T\_PATIENT\_ALLERGY a  JOIN T\_PATIENT b ON a.PATIENT\_ID = b.PATIENT\_ID  WHERE b.RACE = 'asian'  GROUP BY a.ALLERGY\_DESCRIPTION  ORDER BY a.ALLERGY\_DESCRIPTION) h ON a.ALLERGY\_DESCRIPTION = h.ALLERGY\_DESCRIPTION  JOIN (SELECT a.ALLERGY\_DESCRIPTION, COUNT(\*) AS NATIVE\_COUNT  FROM T\_PATIENT\_ALLERGY a  JOIN T\_PATIENT b ON a.PATIENT\_ID = b.PATIENT\_ID  WHERE b.RACE = 'native'  GROUP BY a.ALLERGY\_DESCRIPTION  ORDER BY a.ALLERGY\_DESCRIPTION) i ON a.ALLERGY\_DESCRIPTION = i.ALLERGY\_DESCRIPTION  JOIN (SELECT a.ALLERGY\_DESCRIPTION, COUNT(\*) AS HISPANIC\_COUNT  FROM T\_PATIENT\_ALLERGY a  JOIN T\_PATIENT b ON a.PATIENT\_ID = b.PATIENT\_ID  WHERE b.ETHNICITY = 'hispanic'  GROUP BY a.ALLERGY\_DESCRIPTION  ORDER BY a.ALLERGY\_DESCRIPTION) j ON a.ALLERGY\_DESCRIPTION = j.ALLERGY\_DESCRIPTION  JOIN (SELECT a.ALLERGY\_DESCRIPTION, COUNT(\*) AS NONHISPANIC\_COUNT  FROM T\_PATIENT\_ALLERGY a  JOIN T\_PATIENT b ON a.PATIENT\_ID = b.PATIENT\_ID  WHERE b.ETHNICITY = 'nonhispanic'  GROUP BY a.ALLERGY\_DESCRIPTION  ORDER BY a.ALLERGY\_DESCRIPTION) k ON a.ALLERGY\_DESCRIPTION = k.ALLERGY\_DESCRIPTION; |
| T Result: |  |
| Z Query (SQL): | SELECT a.ALLERGY\_DESCRIPTION, a.ALL\_COUNT, b.MALE\_COUNT, c.FEMALE\_COUNT, d.SINGLE\_COUNT,  e.MARRIED\_COUNT, f.WHITE\_COUNT, g.BLACK\_COUNT, h.ASIAN\_COUNT, i.NATIVE\_COUNT,  j.HISPANIC\_COUNT, k.NONHISPANIC\_COUNT  FROM (SELECT a.ALLERGY\_DESCRIPTION, COUNT(\*) AS ALL\_COUNT  FROM Z\_ALLERGY a  JOIN Z\_PATIENT\_ALLERGY b ON a.ALLERGY\_CODE = b.ALLERGY\_CODE  JOIN Z\_PATIENT c ON b.PATIENT\_ID = c.PATIENT\_ID  GROUP BY a.ALLERGY\_DESCRIPTION  ORDER BY a.ALLERGY\_DESCRIPTION) a  JOIN (SELECT a.ALLERGY\_DESCRIPTION, COUNT(\*) AS MALE\_COUNT  FROM Z\_ALLERGY a  JOIN Z\_PATIENT\_ALLERGY b ON a.ALLERGY\_CODE = b.ALLERGY\_CODE  JOIN Z\_PATIENT c ON b.PATIENT\_ID = c.PATIENT\_ID  WHERE c.GENDER\_ID = 'M'  GROUP BY a.ALLERGY\_DESCRIPTION  ORDER BY a.ALLERGY\_DESCRIPTION) b ON a.ALLERGY\_DESCRIPTION = b.ALLERGY\_DESCRIPTION  JOIN (SELECT a.ALLERGY\_DESCRIPTION, COUNT(\*) AS FEMALE\_COUNT  FROM Z\_ALLERGY a  JOIN Z\_PATIENT\_ALLERGY b ON a.ALLERGY\_CODE = b.ALLERGY\_CODE  JOIN Z\_PATIENT c ON b.PATIENT\_ID = c.PATIENT\_ID  WHERE c.GENDER\_ID = 'F'  GROUP BY a.ALLERGY\_DESCRIPTION  ORDER BY a.ALLERGY\_DESCRIPTION) c ON a.ALLERGY\_DESCRIPTION = c.ALLERGY\_DESCRIPTION  JOIN (SELECT a.ALLERGY\_DESCRIPTION, COUNT(\*) AS SINGLE\_COUNT  FROM Z\_ALLERGY a  JOIN Z\_PATIENT\_ALLERGY b ON a.ALLERGY\_CODE = b.ALLERGY\_CODE  JOIN Z\_PATIENT c ON b.PATIENT\_ID = c.PATIENT\_ID  WHERE c.MARITAL\_STATUS\_ID = 'S'  GROUP BY a.ALLERGY\_DESCRIPTION  ORDER BY a.ALLERGY\_DESCRIPTION) d ON a.ALLERGY\_DESCRIPTION = d.ALLERGY\_DESCRIPTION  JOIN (SELECT a.ALLERGY\_DESCRIPTION, COUNT(\*) AS MARRIED\_COUNT  FROM Z\_ALLERGY a  JOIN Z\_PATIENT\_ALLERGY b ON a.ALLERGY\_CODE = b.ALLERGY\_CODE  JOIN Z\_PATIENT c ON b.PATIENT\_ID = c.PATIENT\_ID  WHERE c.MARITAL\_STATUS\_ID = 'M'  GROUP BY a.ALLERGY\_DESCRIPTION  ORDER BY a.ALLERGY\_DESCRIPTION) e ON a.ALLERGY\_DESCRIPTION = e.ALLERGY\_DESCRIPTION  JOIN (SELECT a.ALLERGY\_DESCRIPTION, COUNT(\*) AS WHITE\_COUNT  FROM Z\_ALLERGY a  JOIN Z\_PATIENT\_ALLERGY b ON a.ALLERGY\_CODE = b.ALLERGY\_CODE  JOIN Z\_PATIENT c ON b.PATIENT\_ID = c.PATIENT\_ID  WHERE c.RACE\_ID = 2  GROUP BY a.ALLERGY\_DESCRIPTION  ORDER BY a.ALLERGY\_DESCRIPTION) f ON a.ALLERGY\_DESCRIPTION = f.ALLERGY\_DESCRIPTION  JOIN (SELECT a.ALLERGY\_DESCRIPTION, COUNT(\*) AS BLACK\_COUNT  FROM Z\_ALLERGY a  JOIN Z\_PATIENT\_ALLERGY b ON a.ALLERGY\_CODE = b.ALLERGY\_CODE  JOIN Z\_PATIENT c ON b.PATIENT\_ID = c.PATIENT\_ID  WHERE c.RACE\_ID = 5  GROUP BY a.ALLERGY\_DESCRIPTION  ORDER BY a.ALLERGY\_DESCRIPTION) g ON a.ALLERGY\_DESCRIPTION = g.ALLERGY\_DESCRIPTION  JOIN (SELECT a.ALLERGY\_DESCRIPTION, COUNT(\*) AS ASIAN\_COUNT  FROM Z\_ALLERGY a  JOIN Z\_PATIENT\_ALLERGY b ON a.ALLERGY\_CODE = b.ALLERGY\_CODE  JOIN Z\_PATIENT c ON b.PATIENT\_ID = c.PATIENT\_ID  WHERE c.RACE\_ID = 1  GROUP BY a.ALLERGY\_DESCRIPTION  ORDER BY a.ALLERGY\_DESCRIPTION) h ON a.ALLERGY\_DESCRIPTION = h.ALLERGY\_DESCRIPTION  JOIN (SELECT a.ALLERGY\_DESCRIPTION, COUNT(\*) AS NATIVE\_COUNT  FROM Z\_ALLERGY a  JOIN Z\_PATIENT\_ALLERGY b ON a.ALLERGY\_CODE = b.ALLERGY\_CODE  JOIN Z\_PATIENT c ON b.PATIENT\_ID = c.PATIENT\_ID  WHERE c.RACE\_ID = 4  GROUP BY a.ALLERGY\_DESCRIPTION  ORDER BY a.ALLERGY\_DESCRIPTION) i ON a.ALLERGY\_DESCRIPTION = i.ALLERGY\_DESCRIPTION  JOIN (SELECT a.ALLERGY\_DESCRIPTION, COUNT(\*) AS HISPANIC\_COUNT  FROM Z\_ALLERGY a  JOIN Z\_PATIENT\_ALLERGY b ON a.ALLERGY\_CODE = b.ALLERGY\_CODE  JOIN Z\_PATIENT c ON b.PATIENT\_ID = c.PATIENT\_ID  WHERE c.ETHNICITY\_ID = 2  GROUP BY a.ALLERGY\_DESCRIPTION  ORDER BY a.ALLERGY\_DESCRIPTION) j ON a.ALLERGY\_DESCRIPTION = j.ALLERGY\_DESCRIPTION  JOIN (SELECT a.ALLERGY\_DESCRIPTION, COUNT(\*) AS NONHISPANIC\_COUNT  FROM Z\_ALLERGY a  JOIN Z\_PATIENT\_ALLERGY b ON a.ALLERGY\_CODE = b.ALLERGY\_CODE  JOIN Z\_PATIENT c ON b.PATIENT\_ID = c.PATIENT\_ID  WHERE c.ETHNICITY\_ID = 1  GROUP BY a.ALLERGY\_DESCRIPTION  ORDER BY a.ALLERGY\_DESCRIPTION) k ON a.ALLERGY\_DESCRIPTION = k.ALLERGY\_DESCRIPTION; |
| Z Result: |  |
| Comments | Any differences on the two results?  Our two results are nearly identical. |

|  |  |
| --- | --- |
| Challenge 3: | **Produce a tabular result (from a single SELECT) that shows living cities in Massachusetts (in rows) and race (each race in a column). Each cell must show the proportion (%) of patients in each City/Race affected by any allergy.** |
| T Query (SQL): | SELECT a.CITY, ROUND((b.WHITE\_COUNT / a.ALL\_COUNT)\*100,2) AS WHITE\_PERCENT, ROUND((c.BLACK\_COUNT/ a.ALL\_COUNT)\*100,2) AS BLACK\_PERCENT,  ROUND((d.ASIAN\_COUNT/ a.ALL\_COUNT)\*100,2) AS ASIAN\_PERCENT, ROUND((e.NATIVE\_COUNT/ a.ALL\_COUNT)\*100,2) AS NATIVE\_PERCENT  FROM (SELECT CITY, COUNT(\*) AS ALL\_COUNT  FROM T\_PATIENT  GROUP BY CITY  ORDER BY CITY) a  JOIN (SELECT CITY, COUNT(\*) AS WHITE\_COUNT  FROM T\_PATIENT  WHERE RACE = 'white'  GROUP BY CITY  ORDER BY CITY) b ON a.CITY = b.CITY  JOIN (SELECT CITY, COUNT(\*) AS BLACK\_COUNT  FROM T\_PATIENT  WHERE RACE = 'black'  GROUP BY CITY  ORDER BY CITY) c ON a.CITY = c.CITY  JOIN (SELECT CITY, COUNT(\*) AS ASIAN\_COUNT  FROM T\_PATIENT  WHERE RACE = 'asian'  GROUP BY CITY  ORDER BY CITY) d ON a.CITY = d.CITY  JOIN (SELECT CITY, COUNT(\*) AS NATIVE\_COUNT  FROM T\_PATIENT  WHERE RACE = 'native'  GROUP BY CITY  ORDER BY CITY) e ON a.CITY = e.CITY; |
| T Result: |  |
| Z Query (SQL): | SELECT a.CITY\_NAME, ROUND((b.WHITE\_COUNT / a.ALL\_COUNT)\*100,2) AS WHITE\_PERCENT, ROUND((c.BLACK\_COUNT/ a.ALL\_COUNT)\*100,2) AS BLACK\_PERCENT,  ROUND((d.ASIAN\_COUNT/ a.ALL\_COUNT)\*100,2) AS ASIAN\_PERCENT, ROUND((e.NATIVE\_COUNT/ a.ALL\_COUNT)\*100,2) AS NATIVE\_PERCENT  FROM (SELECT b.CITY\_NAME, COUNT(\*) AS ALL\_COUNT  FROM Z\_PATIENT a  JOIN Z\_CITY b ON a.LIVING\_PLACE\_CITY\_ID = b.CITY\_ID  GROUP BY b.CITY\_NAME  ORDER BY b.CITY\_NAME) a  JOIN (SELECT b.CITY\_NAME, COUNT(\*) AS WHITE\_COUNT  FROM Z\_PATIENT a  JOIN Z\_CITY b ON a.LIVING\_PLACE\_CITY\_ID = b.CITY\_ID  WHERE a.RACE\_ID = 2  GROUP BY b.CITY\_NAME  ORDER BY b.CITY\_NAME) b ON a.CITY\_NAME = b.CITY\_NAME  JOIN (SELECT b.CITY\_NAME, COUNT(\*) AS BLACK\_COUNT  FROM Z\_PATIENT a  JOIN Z\_CITY b ON a.LIVING\_PLACE\_CITY\_ID = b.CITY\_ID  WHERE a.RACE\_ID = 5  GROUP BY b.CITY\_NAME  ORDER BY b.CITY\_NAME) c ON a.CITY\_NAME = c.CITY\_NAME  JOIN (SELECT b.CITY\_NAME, COUNT(\*) AS ASIAN\_COUNT  FROM Z\_PATIENT a  JOIN Z\_CITY b ON a.LIVING\_PLACE\_CITY\_ID = b.CITY\_ID  WHERE a.RACE\_ID = 1  GROUP BY b.CITY\_NAME  ORDER BY b.CITY\_NAME) d ON a.CITY\_NAME = d.CITY\_NAME  JOIN (SELECT b.CITY\_NAME, COUNT(\*) AS NATIVE\_COUNT  FROM Z\_PATIENT a  JOIN Z\_CITY b ON a.LIVING\_PLACE\_CITY\_ID = b.CITY\_ID  WHERE a.RACE\_ID = 4  GROUP BY b.CITY\_NAME  ORDER BY b.CITY\_NAME) e ON a.CITY\_NAME = e.CITY\_NAME; |
| Z Result: |  |
| Comments | Any differences on the two results?  There are no differences on the two results. |

|  |  |
| --- | --- |
| Challenge 4: | **You MUST propose one additional challenging question with queries and results** |
| Challenge: | **Produce a tabular result (from a single SELECT) that shows living cities (in rows) and ethnicity (each ethnicity in a column).** |
| T Query (SQL): | SELECT a.CITY, ROUND((b.HISPANIC\_COUNT / a.ALL\_COUNT)\*100,2) AS HISPANIC\_PERCENT,  ROUND((c.NONHISPANIC\_COUNT/ a.ALL\_COUNT)\*100,2) AS NONHISPANIC\_PERCENT  FROM (SELECT CITY, COUNT(\*) AS ALL\_COUNT  FROM T\_PATIENT  GROUP BY CITY  ORDER BY CITY) a  JOIN (SELECT CITY, COUNT(\*) AS HISPANIC\_COUNT  FROM T\_PATIENT  WHERE ETHNICITY = 'hispanic'  GROUP BY CITY  ORDER BY CITY) b ON a.CITY = b.CITY  JOIN (SELECT CITY, COUNT(\*) AS NONHISPANIC\_COUNT  FROM T\_PATIENT  WHERE ETHNICITY = 'nonhispanic'  GROUP BY CITY  ORDER BY CITY) c ON a.CITY = c.CITY; |
| T Result: |  |
| Z Query (SQL): | SELECT a.CITY\_NAME, ROUND((b.HISPANIC\_COUNT / a.ALL\_COUNT)\*100,2) AS HISPANIC\_PERCENT,  ROUND((c.NONHISPANIC\_COUNT/ a.ALL\_COUNT)\*100,2) AS NONHISPANIC\_PERCENT  FROM (SELECT b.CITY\_NAME, COUNT(\*) AS ALL\_COUNT  FROM Z\_PATIENT a  JOIN Z\_CITY b ON a.LIVING\_PLACE\_CITY\_ID = b.CITY\_ID  GROUP BY b.CITY\_NAME  ORDER BY b.CITY\_NAME) a  JOIN (SELECT b.CITY\_NAME, COUNT(\*) AS HISPANIC\_COUNT  FROM Z\_PATIENT a  JOIN Z\_CITY b ON a.LIVING\_PLACE\_CITY\_ID = b.CITY\_ID  WHERE a.ETHNICITY\_ID = 2  GROUP BY b.CITY\_NAME  ORDER BY b.CITY\_NAME) b ON a.CITY\_NAME = b.CITY\_NAME  JOIN (SELECT b.CITY\_NAME, COUNT(\*) AS NONHISPANIC\_COUNT  FROM Z\_PATIENT a  JOIN Z\_CITY b ON a.LIVING\_PLACE\_CITY\_ID = b.CITY\_ID  WHERE a.ETHNICITY\_ID = 1  GROUP BY b.CITY\_NAME  ORDER BY b.CITY\_NAME) c ON a.CITY\_NAME = c.CITY\_NAME; |
| Z Result: |  |
| Comments | There are no differences on the two results. |