**Problem Set 3: Advanced SQL Practice I**

*For this problem set, you will use the OMOP database on EduHeLx.*

Answer each of the following questions. For each question, provide the SQL query or queries you used to arrive at your answer.

**1.** Which patient (or patients, if there is a tie) has the longest number of days between their earliest visit and their most recent visit? Output both the person\_id and number of days in the result.

**select p.person\_id, max(a.visit\_end\_date) - min(a.visit\_start\_date) as days\_between**

**from all\_visits a, person p**

**where a.patient = p.person\_source\_value**

**group by p.person\_id**

**order by days\_between desc**

**limit 1;**

**2.** Use a set operator in a query against the *condition\_occurrence* and *procedure\_occurrence* tables to output person\_ids for patients who have ever had nausea (condition\_concept\_id 31967) but have never had a colonoscopy (procedure\_concept\_id 4249893).

**select person\_id**

**from condition\_occurrence**

**where condition\_concept\_id = 31967**

**except**

**select person\_id**

**from procedure\_occurrence**

**where procedure\_concept\_id = 4249;**

**3.** 3025315 is the measurement\_concept\_id for body weight and 3036277 is the measurement\_concept\_id for body height. In this database, weight is measured in kilograms and height is measured in centimeters. Use CASE statements and a single query to convert weight values to pounds and height values to inches and output a table that looks like the following:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **person\_id** | **measurement\_type** | **original\_value** | **original\_unit** | **converted\_value** | **new\_unit** |
| 12234987 | height | 204 | cm | 80.3 | in |
| 12234987 | weight | 54 | kg | 119.1 | lbs |
| 57283992 | weight | 72 | kg | 158.7 | lbs |

Some of these columns are just renamed versions of columns already in the database; other columns will be created from your CASE statements, and do not currently exist in the database.

**select person\_id,**

**case**

**when measurement\_concept\_id = 3025315 then 'weight'**

**when measurement\_concept\_id = 3036277 then 'height'**

**end as measurement\_type,**

**value\_as\_number as original\_value,**

**case**

**when measurement\_concept\_id = 3025315 then 'kg'**

**when measurement\_concept\_id = 3036277 then 'cm'**

**end as original\_unit,**

**case**

**when measurement\_concept\_id = 3025315 then ROUND(value\_as\_number \* 2.2, 1)**

**when measurement\_concept\_id = 3036277 then ROUND(value\_as\_number \* 0.39, 1)**

**end as converted\_value,**

**case**

**when measurement\_concept\_id = 3025315 then 'lbs'**

**when measurement\_concept\_id = 3036277 then 'in'**

**end as new\_unit**

**from measurement**

**where measurement\_concept\_id in (3036277, 3025315);**

**4.** For each patient in the *measurement* table, output the difference between their earliest measured weight and their most recent measured weight, as well as the number of days between those two measurements. Concatenate the unit “kg” onto the weight difference. The resulting table should look like:

|  |  |  |
| --- | --- | --- |
| **person\_id** | **weight\_change** | **days\_between** |
| 2783423 | 8.4 kg | 657 |
| 2978342 | -3.5 kg | 283 |

Do not include any patient with no weight change (i.e., a weight\_change of 0.0) in the final result.

Hint: Getting this output requires multiple subqueries.

**This would be much easier using with…**

**select person\_id,**

**concat((select value\_as\_number**

**from measurement m1**

**where m1.person\_id = m.person\_id**

**and measurement\_concept\_id = 3025315**

**order by measurement\_date desc limit 1) -**

**(select value\_as\_number**

**from measurement m2**

**where m2.person\_id = m.person\_id**

**and measurement\_concept\_id = 3025315**

**order by measurement\_date ASC LIMIT 1), ' kg') as weight\_change,**

**((select measurement\_date**

**from measurement m1**

**where m1.person\_id = m.person\_id**

**and measurement\_concept\_id = 3025315**

**order by measurement\_date desc limit 1) -**

**(select measurement\_date**

**from measurement m2**

**where m2.person\_id = m.person\_id**

**AND measurement\_concept\_id = 3025315**

**order by measurement\_date asc limit 1)) as days\_between**

**from measurement m**

**where measurement\_concept\_id = 3025315**

**group by person\_id**

**having ((select value\_as\_number**

**from measurement m1**

**where m1.person\_id = m.person\_id**

**and measurement\_concept\_id = 3025315**

**order by measurement\_date desc**

**limit 1)-**

**(select value\_as\_number**

**from measurement m2**

**where m2.person\_id = m.person\_id**

**and measurement\_concept\_id = 3025315**

**order by measurement\_date asc limit 1)) <> 0.0;**

**5.** Write SQL queries to identify the following string patterns in the concept\_name field of the concept\_table. Some answers will require regular expressions.

**a.** Select all concept names longer than 250 characters.

**select concept\_name**

**from concept**

**where length(concept\_name) > 250;**

**b.** Select all concept names, case insensitive, where the name contains any one of the following words: head, shoulder, knee, toe.

**select concept\_name**

**from concept**

**where lower(concept\_name) like '%head%'**

**or lower(concept\_name) like '%shoulder%'**

**or lower(concept\_name) like '%knee%'**

**or lower(concept\_name) like '%toe%';**

**c**. For all concept names in (b), replace any instance of head, shoulder, knee, or toe with your first name. Output the version of the concept names with the replacements.

**select regexp\_replace(lower(concept\_name), '(head|shoulder|knee|toe)', 'skylar') as skylar\_concept\_name**

**from concept**

**where lower(concept\_name) like '%head%'**

**or lower(concept\_name) like '%shoulder%'**

**or lower(concept\_name) like '%knee%'**

**or lower(concept\_name) like '%toe%';**

**d.** Select all concept names that contain at least one digit, followed by at least one letter, followed by at least one digit. (E.g., the phrase “4 by 4 inch” would match; the phrase “4 inch heel” would not.)

**select concept\_name**

**from concept**

**where concept\_name ~\* '[0-9]+[a-z]+[0-9]+';**