**CO13:** Distribution of condition start dates

This query is used to to provide summary statistics for condition occurrence start dates (condition\_occurrence\_start\_date) across all condition occurrence records: the mean, the standard deviation, the minimum, the 25th percentile, the median, the 75th percentile, the maximum and the number of missing values. No input is required for this query.

**Sample query:**

with end\_rank as (

SELECT

condition\_start\_date-'0001-01-01' as num\_start\_date,

condition\_start\_date,

sum(1) over (partition by 1 order by condition\_start\_date asc rows between unbounded preceding and current row) as rownumasc

FROM

condition\_occurrence

),

other\_stat as (

SELECT

count(condition\_start\_date) as condition\_start\_date\_count,

min(condition\_start\_date) as condition\_start\_date\_min,

max(condition\_start\_date) as condition\_start\_date\_max,

to\_date('0001-01-01', 'yyyy/mm/dd')+ cast(avg(condition\_start\_date-'0001-01-01') as int) as condition\_start\_date\_average,

stddev((condition\_start\_date-'0001-01-01')) as condition\_start\_date\_stddev

FROM

condition\_occurrence

WHERE

condition\_start\_date is not null

)

SELECT

(SELECT count(condition\_start\_date) FROM condition\_occurrence WHERE condition\_start\_date is null) AS condition\_start\_date\_null\_count,

\*

FROM

other\_stat,

( SELECT

to\_date('0001-01-01', 'yyyy/mm/dd')+cast(avg(condition\_start\_date-'0001-01-01') as int) AS condition\_start\_date\_25percentile

FROM

(select \*,(select count(\*) from end\_rank) as rowno from end\_rank)

WHERE

(rownumasc=cast (rowno\*0.25 as int) and mod(rowno\*25,100)=0) or

(rownumasc=cast (rowno\*0.25 as int) and mod(rowno\*25,100)>0) or

(rownumasc=cast (rowno\*0.25 as int)+1 and mod(rowno\*25,100)>0)

) AS condition\_start\_date\_25percentile,

( SELECT

to\_date('0001-01-01', 'yyyy/mm/dd')+cast(avg(condition\_start\_date-'0001-01-01') as int) as condition\_start\_date\_median

FROM

(select \*,(select count(\*) from end\_rank) as rowno from end\_rank)

WHERE

(rownumasc=cast (rowno\*0.50 as int) and mod(rowno\*50,100)=0) or

(rownumasc=cast (rowno\*0.50 as int) and mod(rowno\*50,100)>0) or

(rownumasc=cast (rowno\*0.50 as int)+1 and mod(rowno\*50,100)>0)

) AS condition\_start\_date\_median,

( SELECT

to\_date('0001-01-01', 'yyyy/mm/dd')+cast(avg(condition\_start\_date-'0001-01-01') as int) as condition\_start\_date\_75percentile

FROM

(select \*,(select count(\*) from end\_rank) as rowno from end\_rank)

WHERE

(rownumasc=cast (rowno\*0.75 as int) and mod(rowno\*75,100)=0) or

(rownumasc=cast (rowno\*0.75 as int) and mod(rowno\*75,100)>0) or

(rownumasc=cast (rowno\*0.75 as int)+1 and mod(rowno\*75,100)>0)

) AS condition\_start\_date\_75percentile;

**Input:**

None

**Output:**

|  |  |
| --- | --- |
| **Field** | **Description** |
| condition\_start\_date\_null\_count | Number of condition occurrences where start date is null |
| condition\_start\_date\_count | Number of condition occurrence start dates |
| condition\_start\_date\_min | The earliest start date of a condition occurrence |
| condition\_start\_date\_max | The latest start date of a condition occurrence |
| condition\_start\_date\_average | The average start date (spanning from the earliest to the latest date and counted by days) |
| condition\_start\_date\_stddev | The standard deviation of start dates, in number of days (spanning from the earliest to the latest date and counted by days) |
| condition\_start\_date\_25percentile | A start date where 25 percent of the other end dates are earlier |
| condition\_start\_date\_median | A start date where half of the other end dates are earlier and half are later |
| condition\_start\_date\_75percentile | A start date where 75 percent of the other end dates are earlier |

**Sample output record:**

|  |  |
| --- | --- |
| **Field** | **Value** |
| condition\_end\_date\_null\_count | 0 |
| condition\_end\_date\_count | 224523674 |
| condition\_end\_date\_min | 2003-01-01 |
| condition\_end\_date\_max | 2011-11-08 |
| condition\_end\_date\_average | 2008-11-30 |
| condition\_end\_date\_stddev | 651.27 |
| condition\_end\_date\_25percentile | 2007-10-30 |
| condition\_end\_date\_median | 2009-05-06 |
| condition\_end\_date\_75percentile | 2010-05-03 |