-- icnarc\_\_02\_\_format\_times

-- This script creates 'icnarc\_02e\_dates\_and\_times\_formatted', a copy of the source table where the dates and times are correctly formatted

-- Note that tables are created with the naming convention

-- icnarc\_XXy\_description, where

-- XX is the number, e.g. '01' relating to the script number

-- y is a letter, e.g. 'a' relating to the order of table creation within the script

-- Tables are dropped when no longer needed

-- Steps:

-- 1) Change any null values to actual NULLs

-- 2) Cast new times

-- In this dataset, times are in the format of a DATETIME, with the default date 1899-12-30. I will be dealing with the times on a case by case basis and have discussed each of these decisions with John.

-- Time values:

-- A) Time\_1st\_managed\_by\_unit\_team - there are 15992 with 00:00:00. The maximum tally for the other distinct values = 15. So it would be fair to assume only around 15 of the 15,992 are truly midnight.

-- Decision: Remove 00:00:00 as a valid time - make all 15,992 null, removing valid data for ~15 values.

-- B) Unit\_admission\_time - Only 6 with 00:00:01, there aren't any 00:00:00. 00:00:01 could be a valid as a time for midnight (Note Discharge\_delay has many times with 1 second and meaningful hour/minute sections).

-- Decision: Change 00:00:01 to 00:00:00 and keep them in.

-- C) Event\_time - Only 3 with 00:00:00, could be valid as midnight.

-- Decision: Keep in.

-- ----- Event\_time removed because no-one knew what it was.

-- D) Discharge\_delay - 6501 with 00:00:00, 436 with 00:00:01. Also discharge delay is a duration not timestamp, so not the best format.

-- Where the delay is greater than 24 hours, the recorded date moves onto the next day(s). Also a high proportion of meaningful hours and minutes have 01 seconds.

-- Decision: i) Remove 00:00:00 - change to null.

-- ii) Remove 1 second where applicable (00:00:01 becomes 00:00:00 and can stay - decided this was valid for a delay of 0).

-- iii) Change to the number of minutes delayed.

-- Use this to get a tally of the unique values

-- SELECT

-- Patient\_discharged,

-- COUNT(Patient\_discharged) AS tally

-- FROM `yhcr-prd-phm-bia-core.CY\_MYSPACE\_EmW.icnarc\_dates\_formatted`

-- GROUP BY Patient\_discharged

-- ORDER BY Patient\_discharged;

-- create empty time variables

DECLARE empty\_time, empty\_time2 STRING;

SET empty\_time = '1899-12-30 00:00:00.000';

SET empty\_time2 = '1899-12-30 00:00:01.000';

-- 1) Change values to null where relevant

CREATE TABLE `yhcr-prd-phm-bia-core.CY\_MYSPACE\_EmW.icnarc\_02a\_times\_nulled`

AS

SELECT \*

EXCEPT(

-- These columns are empty (see post-script)

Booked\_Unit\_Admission\_Time,

Outreach\_last\_visit\_time,

Time\_Extubated,

Bed\_ready\_time,

Event\_time ----- excluded near the end of the cleaning process because it turned out no-one knew what it was

)

REPLACE(

-- A) Time\_1st\_managed\_by\_unit\_team - Change '1899-12-30 00:00:00.000' to null

NULLIF(Time\_1st\_managed\_by\_unit\_team, empty\_time) AS Time\_1st\_managed\_by\_unit\_team,

-- B) Unit\_admission\_time - Change '1899-12-30 00:00:01.000' to '1899-12-30 00:00:00.000' and keep them in.

CASE WHEN Unit\_admission\_time = empty\_time2 THEN empty\_time ELSE Unit\_admission\_time END AS Unit\_admission\_time

-- C) Event\_time - Keep as is [later excluded]

),

-- Di) Discharge\_delay - Change '1899-12-30 00:00:00.000' to null

NULLIF(Discharge\_delay, empty\_time) AS Discharge\_delay\_wip

FROM `yhcr-prd-phm-bia-core.CY\_MYSPACE\_EmW.icnarc\_01b\_dates\_formatted`;

-- Delete dates formatted table

drop table `yhcr-prd-phm-bia-core.CY\_MYSPACE\_EmW.icnarc\_01b\_dates\_formatted`;

-- 2) Change the time columns from STRING types to TIME types (DATETIME for Discharge\_delay\_wip for now)

CREATE TABLE `yhcr-prd-phm-bia-core.CY\_MYSPACE\_EmW.icnarc\_02b\_times\_as\_datetimes`

AS

SELECT \*

REPLACE(

TIME(CAST(Time\_1st\_managed\_by\_unit\_team AS DATETIME)) AS Time\_1st\_managed\_by\_unit\_team,

TIME(CAST(Unit\_Admission\_Time AS DATETIME)) AS Unit\_Admission\_Time,

-----TIME(CAST(Event\_time AS DATETIME)) AS Event\_time,

CAST(Discharge\_delay\_wip AS DATETIME) AS Discharge\_delay\_wip

)

FROM `yhcr-prd-phm-bia-core.CY\_MYSPACE\_EmW.icnarc\_02a\_times\_nulled`;

-- Delete all times nulled table

drop table `yhcr-prd-phm-bia-core.CY\_MYSPACE\_EmW.icnarc\_02a\_times\_nulled`;

-- 3) Make final edits to times (Dii and Diii)

-- Dii) Discharge\_delay\_wip - Change any value stating 1 second to 0 seconds

CREATE TABLE `yhcr-prd-phm-bia-core.CY\_MYSPACE\_EmW.icnarc\_02c\_discharge\_delay\_wip`

AS

SELECT \*

REPLACE(

CASE

WHEN

EXTRACT(SECOND FROM Discharge\_delay\_wip) = 1

THEN

DATETIME\_SUB(Discharge\_delay\_wip, INTERVAL 1 SECOND)

ELSE

Discharge\_delay\_wip

END

AS Discharge\_delay\_wip

)

FROM `yhcr-prd-phm-bia-core.CY\_MYSPACE\_EmW.icnarc\_02b\_times\_as\_datetimes`;

-- Delete all times nulled table

drop table `yhcr-prd-phm-bia-core.CY\_MYSPACE\_EmW.icnarc\_02b\_times\_as\_datetimes`;

-- Diii) Discharge\_delay\_wip - Change to minutes

CREATE TABLE `yhcr-prd-phm-bia-core.CY\_MYSPACE\_EmW.icnarc\_02d\_dates\_and\_times\_formatted`

AS

SELECT \*

EXCEPT(Discharge\_delay\_wip),

-- Get the number of days delayed in minutes

(DATETIME\_DIFF(Discharge\_delay\_wip, DATETIME "1899-12-30 00:00:00", DAY) \* 60 \* 24) +

-- Get the number of hours delayed in minutes

(EXTRACT(HOUR FROM Discharge\_delay\_wip) \* 60) +

-- Get the number of minutes delayed in minutes

EXTRACT(MINUTE FROM Discharge\_delay\_wip) AS Discharge\_delay\_as\_minutes

FROM `yhcr-prd-phm-bia-core.CY\_MYSPACE\_EmW.icnarc\_02c\_discharge\_delay\_wip`;

-- Delete discharge delay wip table

drop table `yhcr-prd-phm-bia-core.CY\_MYSPACE\_EmW.icnarc\_02c\_discharge\_delay\_wip`;

-- Now run 'icnarc\_\_03\_\_format\_other\_columns'

--- TIME COLUMNS (saved as datetime)

---- Usable time columns

--Time\_1st\_managed\_by\_unit\_team

--Unit\_admission\_time (there's a distinct value of '1899-12-30 00:00:01.000')

--Discharge\_delay (many have 01 sec)(there's a distinct value of '1899-12-30 00:00:01.000')

---- Unusable time columns

--Booked\_Unit\_Admission\_Time (only 1 distinct non-null; '1899-12-30 16:12:00.000')

--Outreach\_last\_visit\_time (all '1899-12-30 00:00:00.000')

--Time\_Extubated (all '1899-12-30 00:00:00.000')

--Bed\_ready\_time (all '1899-12-30 00:00:00.000')

--Event\_time ----- excluded near the end of the cleaning process because it turned out no-one knew what it was