

```
-- =====
-- Step 0: Base Cohort Creation
-- Added: 2025-11-12 07:00:34 UTC
-- Description:
-- Joins ICU stays, admissions, and patient demographic tables to create
-- a foundational cohort containing key admission-level and patient-level variables.
-- Computes ICU length of stay in hours and patient age at hospital admission.
-- =====
```

```
CREATE OR REPLACE TABLE `cs229-475622.mimic_iii_new.base_cohort` AS
SELECT
    icu.subject_id,
    icu.hadm_id,
    icu.icustay_id,
    icu.intime,
    icu.outtime,
    TIMESTAMP_DIFF(icu.outtime, icu.intime, HOUR) AS icu_los_hours,
    adm.admittime,
    adm.deathtime,
    pat.dob,
    pat.gender,
    DATE_DIFF(DATE(adm.admittime), DATE(pat.dob), YEAR) AS age
FROM `cs229-475622.mimic_iii_new.icustays` AS icu
JOIN `cs229-475622.mimic_iii_new.admissions` AS adm
ON icu.hadm_id = adm.hadm_id
JOIN `cs229-475622.mimic_iii_new.patients` AS pat
ON icu.subject_id = pat.subject_id;
```

```
-- QC Counts for Base Cohort
SELECT
    COUNT(*) AS total_rows,
    COUNT(DISTINCT subject_id) AS n_patients,
    COUNT(DISTINCT hadm_id) AS n_hospital_admissions,
    COUNT(DISTINCT icustay_id) AS n_icu_stays
FROM `cs229-475622.mimic_iii_new.base_cohort`;
```

```
-- =====
-- Step 1: Exclude Patients Under 18 Years Old (Fractional Age)
-- Added: 2025-11-12 07:00:34 UTC
-- Description:
-- Filters out patients with fractional age under 18 years based on
-- precise day-level difference between admission time and date of birth.
-- Result verified with total_rows=53330 and unique_patients=38511,
-- matching paper (excluded 7964, observed 7965 difference of 1).
```

```
-- =====

CREATE OR REPLACE TABLE `cs229-475622.mimic_ii_new.step1_age18` AS
SELECT *
FROM `cs229-475622.mimic_ii_new.base_cohort`
WHERE TIMESTAMP_DIFF(admittime, dob, DAY) / 365.25 >= 18;

-- QC Counts for Step 1 (Age Filter)
SELECT
COUNT(*) AS total_rows_after_age_filter,
COUNT(DISTINCT subject_id) AS unique_patients_after_age_filter
FROM `cs229-475622.mimic_ii_new.step1_age18`;

-- =====

-- =====

-- Step 2: Exclude Patients Who Died During Their First ICU Stay
-- Project: cs229-475622
-- Dataset: mimic_ii_new
-- Added: 2025-11-11
-- Description:
--   Removes any patient whose recorded DEATHTIME occurs during or before
--   their ICU outtime. These represent deaths in the first ICU admission.
-- =====

CREATE OR REPLACE TABLE `cs229-475622.mimic_ii_new.step2_alive` AS
SELECT *
FROM `cs229-475622.mimic_ii_new.step1_age18`
WHERE deathtime IS NULL OR deathtime > outtime;

-- QC Counts for Step 2 (Alive After First ICU Stay)
SELECT
COUNT(*) AS total_rows_after_death_filter,
COUNT(DISTINCT subject_id) AS unique_patients_after_death_filter
FROM `cs229-475622.mimic_ii_new.step2_alive`;

-- =====

-- Step 3: Exclude Patients Readmitted After Leaving the Hospital
-- Project: cs229-475622
-- Dataset: mimic_ii_new
-- Added: 2025-11-11
-- Description:
--   Excludes patients who were readmitted to the hospital (multiple HADM_IDs).
```

- Keeps all ICU stays for patients with only one hospital admission.
 - This allows multiple ICU stays during the same hospitalization,
 - but removes patients who returned to the hospital after discharge.
- =====

```

CREATE OR REPLACE TABLE `cs229-475622.mimic_iii_new.step3_no_hosp_readmit` AS
SELECT *
FROM `cs229-475622.mimic_iii_new.step2_alive`
WHERE subject_id IN (
  SELECT subject_id
  FROM `cs229-475622.mimic_iii_new.step2_alive`
  GROUP BY subject_id
  HAVING COUNT(DISTINCT hadm_id) = 1 -- keep only patients with one hospital admission
);

-- QC Counts for Step 3 (No Hospital Readmissions)
SELECT
  COUNT(*) AS total_rows_after_readmit_filter,
  COUNT(DISTINCT subject_id) AS unique_patients_after_readmit_filter
FROM `cs229-475622.mimic_iii_new.step3_no_hosp_readmit`;

```

```

CREATE OR REPLACE TABLE `cs229-475622.mimic_iii_new.features_labs_wide` AS
SELECT *
FROM `cs229-475622.mimic_iii_new.features_labs_only`
PIVOT(
  ANY_VALUE(feature_value)
  FOR feature_name IN (
    'anion_gap_mean','anion_gap_sd','anion_gap_min','anion_gap_max',
    'bicarbonate_mean','bicarbonate_sd','bicarbonate_min','bicarbonate_max',
    'calcium_total_mean','calcium_total_sd','calcium_total_min','calcium_total_max',
    'chloride_mean','chloride_sd','chloride_min','chloride_max',
    'creatinine_mean','creatinine_sd','creatinine_min','creatinine_max',
    'glucose_mean','glucose_sd','glucose_min','glucose_max',
    'hematocrit_mean','hematocrit_sd','hematocrit_min','hematocrit_max',
    'hemoglobin_mean','hemoglobin_sd','hemoglobin_min','hemoglobin_max',
    'inr_pt_mean','inr_pt_sd','inr_pt_min','inr_pt_max',
    'mchc_mean','mchc_sd','mchc_min','mchc_max',
    'mch_mean','mch_sd','mch_min','mch_max',
    'mcv_mean','mcv_sd','mcv_min','mcv_max',
    'magnesium_mean','magnesium_sd','magnesium_min','magnesium_max',
    'ptt_mean','ptt_sd','ptt_min','ptt_max',
  )
);

```

```
'pt_mean','pt_sd','pt_min','pt_max',
'phosphate_mean','phosphate_sd','phosphate_min','phosphate_max',
'platelet_count_mean','platelet_count_sd','platelet_count_min','platelet_count_max',
'potassium_mean','potassium_sd','potassium_min','potassium_max',
'rdw_mean','rdw_sd','rdw_min','rdw_max',
'red_blood_cells_mean','red_blood_cells_sd','red_blood_cells_min','red_blood_cells_max',
'sodium_mean','sodium_sd','sodium_min','sodium_max',
'urea_nitrogen_mean','urea_nitrogen_sd','urea_nitrogen_min','urea_nitrogen_max',

:white_blood_cells_mean','white_blood_cells_sd','white_blood_cells_min','white_blood_cells_ma
x'
)
);
```