WITH demographic AS (

SELECT

p.subject\_id,

a.hadm\_id,

p.gender,

FLOOR(TIMESTAMP\_DIFF(a.admittime, p.dob, YEAR)) AS age,

a.ethnicity

FROM `physionet-data.mimiciii\_clinical.patients` p

JOIN `physionet-data.mimiciii\_clinical.admissions` a ON p.subject\_id = a.subject\_id

),

comorbidities AS (

SELECT

d.subject\_id,

d.hadm\_id,

MAX(CASE WHEN d.icd9\_code LIKE '250%' THEN 1 ELSE 0 END) AS diabetes,

MAX(CASE WHEN d.icd9\_code LIKE '401%' THEN 1 ELSE 0 END) AS hypertension,

MAX(CASE WHEN d.icd9\_code LIKE '410%' THEN 1 ELSE 0 END) AS myocardial\_infarction,

MAX(CASE WHEN d.icd9\_code LIKE '430%' OR d.icd9\_code LIKE '431%' OR d.icd9\_code LIKE '432%' THEN 1 ELSE 0 END) AS cerebral\_vascular\_disease,

MAX(CASE WHEN d.icd9\_code LIKE '272%' THEN 1 ELSE 0 END) AS hyperlipidemia,

MAX(CASE WHEN d.icd9\_code LIKE '414%' THEN 1 ELSE 0 END) AS coronary\_heart\_disease,

MAX(CASE WHEN d.icd9\_code LIKE '493%' THEN 1 ELSE 0 END) AS asthma,

MAX(CASE WHEN d.icd9\_code LIKE '571%' THEN 1 ELSE 0 END) AS chronic\_liver\_disease,

MAX(CASE WHEN d.icd9\_code LIKE '585%' THEN 1 ELSE 0 END) AS chronic\_renal\_disease,

MAX(CASE WHEN d.icd9\_code LIKE '140%' OR d.icd9\_code LIKE '141%' THEN 1 ELSE 0 END) AS cancer

FROM `physionet-data.mimiciii\_clinical.diagnoses\_icd` d

GROUP BY d.subject\_id, d.hadm\_id

),

vital\_signs AS (

SELECT

ce.subject\_id,

ce.hadm\_id,

MAX(CASE WHEN ce.itemid = 220045 THEN ce.valuenum END) AS heart\_rate,

MAX(CASE WHEN ce.itemid = 220210 THEN ce.valuenum END) AS systolic\_bp,

MAX(CASE WHEN ce.itemid = 220277 THEN ce.valuenum END) AS gcs

FROM `physionet-data.mimiciii\_clinical.chartevents` ce

GROUP BY ce.subject\_id, ce.hadm\_id

),

lab\_tests AS (

SELECT

le.subject\_id,

le.hadm\_id,

MAX(CASE WHEN le.itemid = 51221 THEN le.valuenum END) AS white\_blood\_cell,

MAX(CASE WHEN le.itemid = 51265 THEN le.valuenum END) AS hemoglobin,

MAX(CASE WHEN le.itemid = 51277 THEN le.valuenum END) AS platelet

FROM `physionet-data.mimiciii\_clinical.labevents` le

GROUP BY le.subject\_id, le.hadm\_id

),

procedures AS (

SELECT

pe.subject\_id,

pe.hadm\_id,

MAX(CASE WHEN pe.itemid = 225792 THEN 1 ELSE 0 END) AS tracheostomy,

MAX(CASE WHEN pe.itemid = 220210 THEN 1 ELSE 0 END) AS vasopressor

FROM `physionet-data.mimiciii\_clinical.procedureevents\_mv` pe

GROUP BY pe.subject\_id, pe.hadm\_id

),

outcomes AS (

SELECT

a.subject\_id,

a.hadm\_id,

CASE WHEN a.deathtime IS NOT NULL AND TIMESTAMP\_DIFF(a.deathtime, a.admittime, DAY) <= 30 THEN 1 ELSE 0 END AS mortality\_30d,

TIMESTAMP\_DIFF(a.dischtime, a.admittime, DAY) AS hospital\_stay\_length,

icu.icustay\_length

FROM `physionet-data.mimiciii\_clinical.admissions` a

LEFT JOIN (

SELECT

icu.subject\_id,

icu.hadm\_id,

TIMESTAMP\_DIFF(icu.outtime, icu.intime, DAY) AS icustay\_length

FROM `physionet-data.mimiciii\_clinical.icustays` icu

) icu ON a.subject\_id = icu.subject\_id AND a.hadm\_id = icu.hadm\_id

),

comorbidity\_percentages AS (

SELECT

COUNT(CASE WHEN com.diabetes = 1 THEN 1 END) / COUNT(\*) \* 100 AS diabetes\_percentage,

COUNT(CASE WHEN com.hypertension = 1 THEN 1 END) / COUNT(\*) \* 100 AS hypertension\_percentage,

COUNT(CASE WHEN com.myocardial\_infarction = 1 THEN 1 END) / COUNT(\*) \* 100 AS myocardial\_infarction\_percentage,

COUNT(CASE WHEN com.cerebral\_vascular\_disease = 1 THEN 1 END) / COUNT(\*) \* 100 AS cerebral\_vascular\_disease\_percentage,

COUNT(CASE WHEN com.hyperlipidemia = 1 THEN 1 END) / COUNT(\*) \* 100 AS hyperlipidemia\_percentage,

COUNT(CASE WHEN com.coronary\_heart\_disease = 1 THEN 1 END) / COUNT(\*) \* 100 AS coronary\_heart\_disease\_percentage,

COUNT(CASE WHEN com.asthma = 1 THEN 1 END) / COUNT(\*) \* 100 AS asthma\_percentage,

COUNT(CASE WHEN com.chronic\_liver\_disease = 1 THEN 1 END) / COUNT(\*) \* 100 AS chronic\_liver\_disease\_percentage,

COUNT(CASE WHEN com.chronic\_renal\_disease = 1 THEN 1 END) / COUNT(\*) \* 100 AS chronic\_renal\_disease\_percentage,

COUNT(CASE WHEN com.cancer = 1 THEN 1 END) / COUNT(\*) \* 100 AS cancer\_percentage

FROM comorbidities com

)

SELECT

demo.subject\_id,

demo.hadm\_id,

demo.gender,

demo.age,

demo.ethnicity,

com.diabetes,

com.hypertension,

com.myocardial\_infarction,

com.cerebral\_vascular\_disease,

com.hyperlipidemia,

com.coronary\_heart\_disease,

com.asthma,

com.chronic\_liver\_disease,

com.chronic\_renal\_disease,

com.cancer,

vs.heart\_rate,

vs.systolic\_bp,

vs.gcs,

lab.white\_blood\_cell,

lab.hemoglobin,

lab.platelet,

proc.tracheostomy,

proc.vasopressor,

out.mortality\_30d,

out.icustay\_length,

out.hospital\_stay\_length,

perc.diabetes\_percentage,

perc.hypertension\_percentage,

perc.myocardial\_infarction\_percentage,

perc.cerebral\_vascular\_disease\_percentage,

perc.hyperlipidemia\_percentage,

perc.coronary\_heart\_disease\_percentage,

perc.asthma\_percentage,

perc.chronic\_liver\_disease\_percentage,

perc.chronic\_renal\_disease\_percentage,

perc.cancer\_percentage

FROM demographic demo

LEFT JOIN comorbidities com ON demo.subject\_id = com.subject\_id AND demo.hadm\_id = com.hadm\_id

LEFT JOIN vital\_signs vs ON demo.subject\_id = vs.subject\_id AND demo.hadm\_id = vs.hadm\_id

LEFT JOIN lab\_tests lab ON demo.subject\_id = lab.subject\_id AND demo.hadm\_id = lab.hadm\_id

LEFT JOIN procedures proc ON demo.subject\_id = proc.subject\_id AND demo.hadm\_id = proc.hadm\_id

LEFT JOIN outcomes out ON demo.subject\_id = out.subject\_id AND demo.hadm\_id = out.hadm\_id

LEFT JOIN comorbidity\_percentages perc ON 1 = 1; -- This adds the percentage data for comorbidities

**THE QUERY for the Target variable “VAP”**

WITH tbi\_patients AS (

SELECT DISTINCT p.subject\_id, a.hadm\_id

FROM `physionet-data.mimiciii\_clinical.patients` p

JOIN `physionet-data.mimiciii\_clinical.admissions` a ON p.subject\_id = a.subject\_id

JOIN `physionet-data.mimiciii\_clinical.diagnoses\_icd` d ON a.hadm\_id = d.hadm\_id

WHERE

(d.icd9\_code BETWEEN '80000' AND '80199') OR

(d.icd9\_code BETWEEN '80300' AND '80499') OR

(d.icd9\_code BETWEEN '8500' AND '85419') -- TBI ICD-9 Codes

),

vap\_patients AS (

SELECT DISTINCT subject\_id, hadm\_id, 1 AS vap

FROM `physionet-data.mimiciii\_clinical.diagnoses\_icd`

WHERE icd9\_code = '99731' -- VAP ICD-9 Code

)

-- Combine the two to add VAP status for TBI patients

SELECT

t.subject\_id,

t.hadm\_id,

COALESCE(v.vap, 0) AS vap -- 1 if VAP is present, 0 if not

FROM tbi\_patients t

LEFT JOIN vap\_patients v ON t.subject\_id = v.subject\_id AND t.hadm\_id = v.hadm\_id

ORDER BY t.subject\_id, t.hadm\_id;